GEORGIA INSTITUTE OF TECHNOLOGY — OFFICE OF CONTRACT ADMINISTRATION

PROJECT ADMINISTRATION DATA SHEET

[X] ORIGINAL
[ ] REVISION NO. ____

Project No. E-21-659 DATE: 4/13/81

Project Director: M. G. Moharam School/Lab Electrical Engineering

Sponsor: National Science Foundation, Washington, D. C. 20550

Type Agreement: Grant No. ECS-8105483

Award Period: From 7/1/81 To 12/31/83* (Performance) - - - - - - - - (Reports)

Sponsor Amount: $48,000 Contracted through:

Cost Sharing: $ 8,585 (E-21-363) GTRI/XXX

Title: "Research Initiation: A Study of Planar Dielectric Grating Couplers"

ADMINISTRATIVE DATA OCA CONTACT Leamon R. Scott, Ext. 4820

1) Sponsor Technical Contact: C. Paul Christensen, NSF Program Official; National Science Foundation; Washington, D. C. 20550 Phone: 202-357-9618

2) Sponsor Admin./Contractual Contact: Lee A. DeHerrera, NSF Grants Official; National Science Foundation; Washington, D. C. 20550 Phone: 202-357-9602

Reports: See Deliverable Schedule Security Classification: N/A

Defense Priority Rating: N/A

RESTRICTIONS

See Attached NSF Supplemental Information Sheet for Additional Requirements.

Travel: Foreign travel must have prior approval - Contact OCA in each case. Domestic travel requires sponsor approval where total will exceed greater of $500 or 125% of approved proposal budget category.

Equipment: Title vests with Georgia Institute of Technology

COMMENTS: *Includes the usual six (6) month unfunded flexibility period.

COPIES TO:

Administrative Coordinator
Research Property Management
Accounting Office
Procurement Office
Research Security Services
Reports Coordinator (OCA)
Legal Services (OCA)
Library, Technical Reports
EES Information Office (2)
Project File (OCA)
Other: ____

GEORGIA INSTITUTE OF TECHNOLOGY OFFICE OF CONTRACT ADMINISTRATION

SPONSORED PROJECT TERMINATION/CLOSEOUT SHEET

Date 1/18/85

Project No. E-21-659 School/~~Lab~~ EE

Includes Subproject No.(s)

Project Director(s) M. G. Moharam GTRI / ~~GIT~~

Sponsor National Science Foundation

Title "Research Initiation: A Study of Planar Dielectric Grating Couplers"

Effective Completion Date: 12/31/83 (Performance) 3/31/84 (Reports)

Grant/Contract Closeout Actions Remaining:

- [ ] None
- [X] Final Invoice or Final Fiscal Report FCTR
- [ ] Closing Documents
- [X] Final Report of Inventions
- [ ] Govt. Property Inventory & Related Certificate
- [ ] Classified Material Certificate
- [ ] Other

Continues Project No. Continued by Project No.

COPIES TO:

Project Director
Research Administrative Network
Research Property Management
Accounting
Procurement/EES Supply Services
~~Research Security Services~~
Reports Coordinator (OCA)
Library
GTRI
Research Communications (2)
Project File
Other A. Jones; M. Heyser

E-21-659

GEORGIA INSTITUTE OF TECHNOLOGY
SCHOOL OF ELECTRICAL ENGINEERING
ATLANTA, GEORGIA 30332

TELEPHONE: (404) 894-2907

September 22, 1982

Dr. James C. Aller
Program Director for Quantum
Electronics, Waves and Beams
National Science Foundation
Washington, DC 20550

SUBJECT: Annual Progress Report for NSF Grant No. ECS-8105483
"A Study of Planar Dielectric Grating Couplers"
(Covering 1 July 1981 to 30 June 1982)

Dear Dr. Aller:

Significant progress has been made during the first year of the above research grant.

For the first time, a state variable approach from linear system theory has been applied by us to solve the exact electromagnetic boundary value problem associated with general slanted planar grating diffraction. This has been done in rigorously and has resulted in a method of solution easily implemented and without approximations. With this very powerful and exact method of analysis, which is applicable to TE and TM polarizations, to lossy and to mixed gratings, previously used approximations have been evaluated for the first time. This new method has been duplicated in several universities, government laboratories, and in high technology companies. These results are reported in:

> Moharam, M. G. and Gaylord, T. K., "Rigorous coupled-wave analysis of planar gratings," J. Opt. Soc. Am., vol. 71, pp. 811-818, July 1981.
>
> Moharam, M. G. and Gaylord, T. K., "Coupled-wave analysis of reflection gratings," Applied Optics, vol. 20, pp. 240-244, January 1981.

A new chain-matrix based method to analyze arbitrary-thickness pure reflection gratings has been developed. For the first time, this exact, rigorous, and simple method has been used to analyze lossy pure reflection gratings with non-sinusoidal modulation for both TE and TM polarizations. This work is reported in:

> Moharam, M. G. and Gaylord, T. K., "Rigorous chain matrix analysis of arbitrary-thickness dielectric reflection gratings," J. Opt. Soc. Am., vol. 72, pp. 187-190, February 1982.

For the first time surface-relief grating has been rigorously analyzed using a coupled-wave approach. The exact method allows the evaluation of the various approximations used in previous work. Surface-relief type gratings are of great interest in quantum electronics, spectroscopy, holography and integrated-optics. This work is reported in:

> Moharam, M. G. and Gaylord, T. K., "Diffraction analysis of dielectric surface-relief gratings," J. Opt. Soc. Am., vol. 72, pp. 000-000, October 1982.

As a result of our research work an invited paper reviewing the different theories used to analyze diffraction gratings and a second paper clarifying the different terminologies used in grating diffraction have been published. This work is reported in:

> Gaylord, T. K. and Moharam, M. G., "Planar dielectric grating diffraction theories," Applied Physics B, vol. 28, pp. 1-14, May 1982. (invited)
>
> Gaylord, T. K. and Moharam, M. G., "Thin and Thick gratings: terminology clarification," Applied Optics, vol. 20, pp. 3271-3273, October 1981.

We are continuing investigating planar and surface grating diffraction for application to beam-couplers. Our plans for the next year are to study the coupling characteristics and performance of grating-couplers.

If additional information is needed, please contact me and I will supply it to you. Your support is deeply appreciated. Thank you.

Sincerely,

M. G. Moharam
Assistant Professor

MGM:svs

Enclosures: Two copies each of above cited papers.

Rigorous coupled-wave analysis of planar-grating diffraction

M. G. Moharam and T. K. Gaylord

School of Electrical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

Received July 7, 1980; revised manuscript received February 18, 1981

A rigorous coupled-wave approach is used to analyze diffraction by general planar gratings bounded by two different media. The grating fringes may have any orientation (slanted or unslanted) with respect to the grating surfaces. The analysis is based on a state-variables representation and results in a unifying, easily computer-implementable matrix formulation of the general planar-grating diffraction problem. Accurate diffraction characteristics are presented for the first time to the authors' knowledge for general slanted gratings. This present rigorous formulation is compared with rigorous modal theory, approximate two-wave modal theory, approximate multiwave coupled-wave theory, and approximate two-wave coupled-wave theory. Typical errors in the diffraction characteristics introduced by these various approximate theories are evaluated for transmission, slanted, and reflection gratings. Inclusion of higher-order waves in a theory is important for obtaining accurate predictions when forward-diffracted orders are dominant (transmission-grating behavior). Conversely, when backward-diffracted orders dominate (reflection-grating behavior), second derivatives of the field amplitudes and boundary diffraction need to be included to produce accurate results.

1. INTRODUCTION

The diffraction of electromagnetic waves by planar gratings has been extensively studied in recent years. These periodic structures have applications in several diverse areas, such as acousto-optics, integrated optics, holography, and spectroscopy. Several different techniques have been used to analyze the diffraction of electromagnetic waves by spatially modulated media. The most common of these methods are the coupled-wave approach[1–7] and the modal approach.[8–17] Both of these approaches are capable of being formulated in a rigorous manner without approximations. As such they have been shown to be exact and equivalent.[18] However, until the present work, accurate calculations have been rather difficult and time consuming. This paper presents a straightforward rigorous coupled-wave analysis that is easily implemented on a computer. The method of solution is based on a state-space representation of the governing set of differential equations.

The coupled-wave approach has long been known to offer superior physical insight into wave-diffraction phenomena, and it frequently yields simple analytical results in limiting cases. In this approach several assumptions typically are made in order to obtain solutions. These approximations are (1) neglecting boundary diffraction,[1–5,7] (2) neglecting the second derivatives of the field amplitudes,[1–5,7] and (3) retaining only one diffracted wave[1,2,4,6] (in addition to the transmitted wave). The first two of these approximations are interconnected and cannot be separated for the case of a planar grating. The full modal approach is a rigorous exact analysis, but solutions have been published only for the unslanted pure transmission case in which the grating vector is parallel to the grating surface. Several authors have attempted to analyze general slanted gratings. Kogelnik[4] and Magnusson and Gaylord[7] applied the coupled-wave approach, with various assumptions discussed above, to this problem. Clearly these analyses are not exact, and in some cases they give obviously incorrect results because of the use of these approximations. Bergstein and Kermisch[12] applied the modal approach to slanted gratings, but they assumed a two-wave regime to obtain solutions. Chu and Kong[17] have also presented a generalized modal approach formulation for the general slanted-gratings case. However, they did not present any calculated results for general slanted gratings because the formulation for slanted gratings in the modal approach results in a complicated transcendental relationship that may be difficult to solve in general. This paper presents, for the first time to the authors' knowledge, accurate diffraction results for slanted gratings.

By using the straightforward rigorous method of analysis presented in this paper, it is a simple matter to analyze the precise effect of using the various approximations mentioned above. The effect of neglecting boundary diffraction and second derivatives of field amplitudes is studied by comparing the present theory with multiwave coupled-wave theory.[7] The effect of neglecting higher-order waves (retaining only one diffracted wave) is analyzed by comparing the present theory and the two-wave modal theory.[12] The combined effect of these approximations is studied by comparing the present theory with the two-wave coupled-wave theory.[4] These three errors are calculated and presented for the first time for several typical transmission, reflection, and general planar gratings.

The rigorous coupled-wave approach presented here analyzes the diffraction of an electromagnetic plane wave incident obliquely at a planar grating bounded by two different media. In general, these gratings simultaneously exhibit both transmission and reflection behavior, as indicated by the forward-diffracted and backward-diffracted waves shown in Fig. 1. This analysis is applicable (1) to holographic gratings in air or other media ($\epsilon_1 = \epsilon_3 \neq \epsilon_2$), (2) to acousto-optic gratings within a medium ($\epsilon_1 = \epsilon_3 = \epsilon_2$), and (3) to grating couplers such as are used in integrated optics ($\epsilon_1 \neq \epsilon_2 \neq \epsilon_3 \neq \epsilon_1$).

Reprinted from Journal of the Optical Society of America,
Vol. 71, page 811, July, 1981

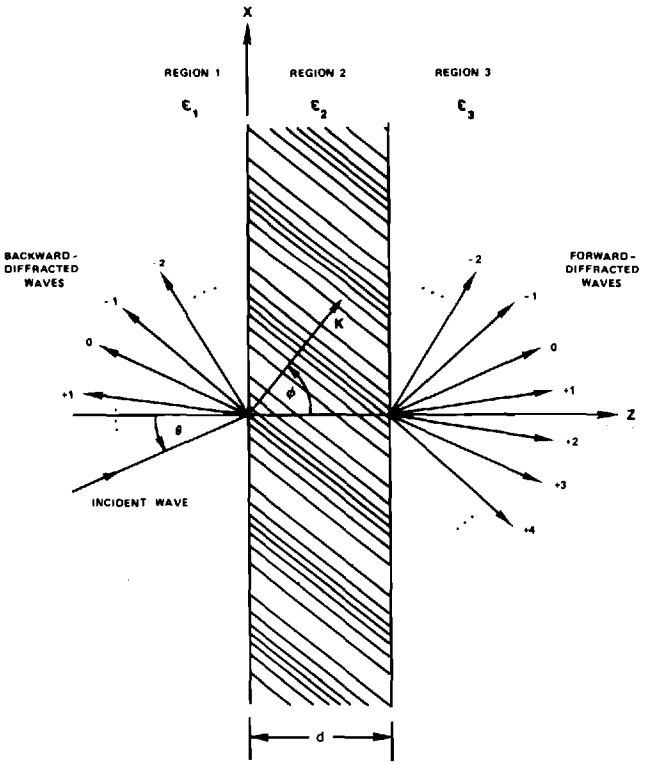


Fig. 1. Geometry for planar-grating diffraction.

2. THEORY

As in most of the work previously cited, the problem under consideration is the diffraction of an obliquely incident plane wave on a lossless (pure phase) sinusoidal grating with the incident wave polarized perpendicular to the plane of incidence (*H* mode). Therefore the electric field will have only one component (in the y direction of Fig. 1). The relative permittivity in the modulated region ($0 < z < d$) is

$$\epsilon(x,z) = \epsilon_2 + \Delta\epsilon \cos[K(x \sin\phi + z \cos\phi)], \quad (1)$$

where ϵ_2 is the average dielectric constant, $\Delta\epsilon$ is the amplitude of the sinusoidal relative permittivity, ϕ is the grating slant angle, and $K = 2\pi/\Lambda$, where Λ is the grating period. The dielectric constant in the unmodulated regions ($z < 0$ and $z > d$) is ϵ_1 and ϵ_3, respectively. It is assumed that each of the three regions has the permeability of free space.

The general approach to the planar-grating problem involves finding a solution of the wave equation in each of the three regions and then matching the tangential electric and magnetic fields at the two interfaces ($z = 0$ and $z = d$) to determine the unknown constants (resulting from solving the differential-wave equation). In region 1, backward-diffracted waves exist. In general, these waves are produced both by diffraction from within the grating volume (bulk diffraction) and by diffraction from the periodic boundary at $z = 0$ (boundary diffraction). These diffraction processes produce a spectrum of plane waves traveling back into region 1 ($z < 0$). The normalized wave amplitudes in region 1 may be expressed as[8-10]

$$E_1 = \exp[-j(\beta_0 x + \xi_{10} z)] + \sum_i R_i \exp[-j(\beta_i x - \xi_{1i} z)], \quad (2)$$

where $\beta_i = k_1 \sin\theta - iK \sin\phi$ for any integer i (the wave index); $\xi_{li}^2 = k_l^2 - \beta_i^2$ for $l = 1, 3$ (the region index); $k_l = 2\pi\epsilon_l^{1/2}/\lambda$ for $l = 1, 2, 3$; λ is the free-space wavelength; $j = (-1)^{1/2}$; θ is the angle of incidence in region 1; and R_i is the normalized amplitude of the ith reflected wave and is to be determined from the matching of the electric and magnetic fields. In region 3 ($z > d$) the spectrum of transmitted plane waves may be expressed as[16,17]

$$E_3 = \sum_i T_i \exp\{-j[\beta_i x + \xi_{3i}(z - d)]\}, \quad (3)$$

where T_i is the normalized amplitude of the ith transmitted wave to be determined from the field matching. In region 2, the modulated region ($0 < z < d$), the electric field may be expressed as

$$E_2 = \sum_i \hat{S}_i(z) \exp[-j(\beta_i x + \xi_{2i} z)], \quad (4)$$

where $\xi_{2i} = k_2 \cos\theta' - iK \cos\phi$, θ' is the angle of refraction inside the modulated region, and $\hat{S}_i(z)$ is the normalized amplitude of the ith wave field at any point within the modulated region. For a given value of i, the wave field inside the grating is not a simple plane wave. It may be expressed as a superposition of an infinite number of plane waves (inherent in the coupled-wave formulation). This superposition includes forward-traveling waves (components in $+z$ direction) and corresponding backward-traveling waves (components in $-z$ direction). These amplitudes are to be determined from solving the modulated-region wave equation

$$\nabla^2 E_2 + (2\pi/\lambda)^2 \epsilon(x,z) E_2 = 0. \quad (5)$$

Note that the three electric fields E_1, E_2, and E_3 in the three regions are phase matched along the two interfaces. It is important to point out that the above analysis is valid for all slant angles ϕ except when the slant angle is identically zero (pure reflection grating). In this case, the modulation is no longer periodic since it has only a finite number of cycles. Equations (2)–(4) are thus not valid because they are derived from the Floquet theorem, which is correct only for infinite periodic structures. However, the above analysis can be used to analyze pure reflection gratings ($\phi = 0$) by considering $\phi = \delta$ and allowing δ to become arbitrarily small. As ϕ approaches (but does not reach) zero, the slab retains a periodicity in the modulation along its boundaries. All the R_i and T_i amplitudes for the reflected and transmitted waves may then be calculated. As ϕ approaches zero, all the reflected wave vectors in region 1 approach a common direction, and the resulting reflected intensity of the composite wave is $\Sigma_i |R_i|^2$. Likewise, all the transmitted wave vectors in region 3 approach a single direction, and the resulting transmitted intensity of this composite wave is $\Sigma_i |T_i|^2$. For a more detailed discussion of planar reflection gratings, see Ref. 19.

3. METHOD OF SOLUTION

To obtain the diffracted amplitudes $\hat{S}_i(z)$, Eqs. (1) and (4) are substituted into Eq. (5), resulting in the infinite set of coupled-wave equations

$$(\Delta\epsilon/8\epsilon_2)\frac{d^2 S_i(u)}{du^2} = (\cos\theta' - i\mu\cos\phi) \times \frac{dS_i(u)}{du} - \rho i(i - B)S_i(u) + S_{i+1}(u) + S_{i-1}(u), \quad (6)$$

with $S_i(u) = \hat{S}_i(z)$, $u = j\pi\Delta\epsilon z/2\lambda(\epsilon_2)^{1/2} = j\kappa z$, $\mu = \lambda/\Lambda(\epsilon_2)^{1/2}$, $\rho = 2\lambda^2/\Lambda^2\Delta\epsilon = 2\mu^2\epsilon_2/\Delta\epsilon$, and $B = 2\Lambda(\epsilon_2)^{1/2}\cos(\phi - \theta')/\lambda = 2\cos(\phi - \theta')/\mu$. This system of coupled-wave equations has been derived *without* the common assumptions and approximations associated with previous coupled-wave analyses,[1–7] such as neglecting the second derivatives of the amplitudes. Therefore the present analysis is as rigorous and as exact as the modal approach. There are five fundamental parameters $[\theta', \phi, \Delta\epsilon/\epsilon_2, d/\Lambda, \lambda/\Lambda(\epsilon_2)^{1/2}]$ contained within Eq. (6). The composite parameters used in Eq. (6) may be expressed in terms of these parameters. Other equivalent sets of parameters obviously may also be chosen. The regime parameter ρ determines the boundary between the Bragg regime and the intermediate diffraction regime.[20] The parameter B represents the Bragg condition, i.e. for incidence at the pth Bragg condition $B = p$ (note that B is not an integer for off-Bragg incidence). The parameter κ is the widely used coupling coefficient. As is shown in the Appendix, if S and S' are defined as state variables, Eq. (6) may be written in matrix form as

$$\begin{bmatrix} \mathbf{S}' \\ \mathbf{S}'' \end{bmatrix} = \begin{bmatrix} b_{rs} \end{bmatrix} \begin{bmatrix} \mathbf{S} \\ \mathbf{S}' \end{bmatrix}, \tag{7}$$

where $\mathbf{S}$, $\mathbf{S}'$, and $\mathbf{S}''$ indicate the column vectors of S_i, dS_i/du, and d^2S_i/du^2, respectively. The quantity $[b]$ is the coefficient matrix specified from Eq. (6). Equation (7) corresponds to an unforced state equation in the state-space description of linear systems. The system of differential equations given by Eq. (7) has a relatively simple and straightforward solution, obtainable in terms of the eigenvalues and the eigenvectors of the coefficient matrix $[b]$. It is

$$S_i(u) = \sum_m C_m w_{im} \exp(q_m u), \tag{8}$$

where q_m is the mth eigenvalue and w_{im} is the mth element of the row in the matrix $[w]$ composed of the eigenvectors, corresponding to the ith wave (not the ith row of the matrix $[w]$). The coefficients C_m are unknown constants to be determined together with R_i and T_i by matching the tangential electric and magnetic fields at the two boundaries ($z = 0$ and $z = d$). The four quantities to be matched and the resulting boundary conditions are tangential E at $z = 0$:

$$R_i + \delta_{i0} = \sum_m C_m w_{im}, \tag{9}$$

tangential H at $z = 0$:

$$\xi_{1i}(R_i - \delta_{i0}) = \sum_m C_m w_{im}(q_m\kappa - \xi_{2i}), \tag{10}$$

tangential E at $z = d$:

$$T_i = \sum_m C_m w_{im} \exp[j(q_m\kappa - \xi_{2i})d], \tag{11}$$

tangential H at $z = d$:

$$-\xi_{3i}T_i = \sum_m C_m w_{im}(q_m\kappa - \xi_{2i}) \exp[j(q_m\kappa - \xi_{2i})d], \tag{12}$$

where δ_{i0} is the Kronecker delta function. Eliminating T_i and R_i from these equations gives

$$2\delta_{i0}\xi_{1i} = \sum_m C_m w_{im}(\xi_{2i} - q_m\kappa + \xi_{1i}), \tag{13}$$

$$0 = \sum_m C_m w_{im}(\xi_{2i} - q_m\kappa - \xi_{3i}) \exp(jq_m\kappa). \tag{14}$$

Note that

$$\xi_{2i} = (2\rho\kappa/\mu^2)(\cos\theta' - i\mu\cos\phi), \tag{15}$$

$$\xi_{li} = (2\rho\kappa/\mu^2)[(\epsilon_l/\epsilon_2) - (\sin\theta' - i\mu\sin\phi)^2]^{1/2} \qquad l = 1, 3. \tag{16}$$

In regions 1 and 3, ξ_{li} is either positive real (propagating wave) or negative imaginary (evanescent wave). The system of linear equations given by Eqs. (13) and (14) can be solved for C_m, and then R_i and T_i can be calculated from Eqs. (9) and (11). Note that the number of equations available is exactly equal to the number of unknowns. For example, if n waves are retained in the analysis, then there will be n unknown values each of R_i and of T_i and $2n$ unknown values of C_m. This is because the coefficient matrix $[b]$ in Eq. (7) is a $2n \times 2n$ matrix and therefore has $2n$ eigenvalues, and thus there are $2n$ unknown values of C_m. Alternatively, this may be viewed as being due to the n coupled-wave equations, each being a second-order differential equation, and thus there are $2n$ roots or eigenvalues and $2n$ unknown constants C_m to be determined from the boundary conditions. Therefore the total number of unknowns is $4n$, and Eqs. (9)–(12) provide $4n$ linear equations in these unknowns.

To summarize, the algorithm used to solve this problem proceeds as follows: First the coefficient matrix $[b]$ is constructed, and then eigenvalues and eigenvectors are calculated (typically by using a computer library program). The system of linear equations, Eqs. (13) and (14), is then constructed and solved for C_m (using a technique such as gauss elimination). Equations (9) and (11) are then used to calculate the diffracted amplitudes R_i and T_i. Power conservation requires that the sum of the efficiencies for all of the propagating waves be unity. That is,

$$\sum_i (DE_{1i} + DE_{3i}) = 1, \tag{17}$$

where DE_{1i} and DE_{3i} are the diffraction efficiencies in regions 1 and 3, respectively. These diffraction efficiencies are given by

$$DE_{1i} = Re(\xi_{1i}/\xi_{10})R_iR_i^* \tag{18}$$

and

$$DE_{3i} = Re(\xi_{3i}/\xi_{10})T_iT_i^*. \tag{19}$$

The real part of the ratio of the propagation constants occurs when the time-average power-flow density is obtained by taking the real part of the complex Poynting vector. The quantity $\mathrm{Re}(\xi_{li}/\xi_{10})$ is just the usual ratio of the cosine of the diffraction angle for the ith wave to the cosine of the angle of the zeroth-order wave for the lth medium. The results of sample calculations for the diffraction efficiencies are shown in Figs. 2–6 as a function of the grating strength parameter γ ($= \kappa d/\cos\theta'$) and as a function of d/Λ for a pure transmission grating ($\phi = \pi/2$), for general slanted gratings ($\phi = \pi/3, \pi/6$), and for a pure reflection grating ($\phi = 0$).

4. COMPARISON WITH PREVIOUS ANALYSES

A. Rigorous Modal Approach

Basically, the rigorous coupled-wave approach presented in this paper and the rigorous modal approach analyze the pla-

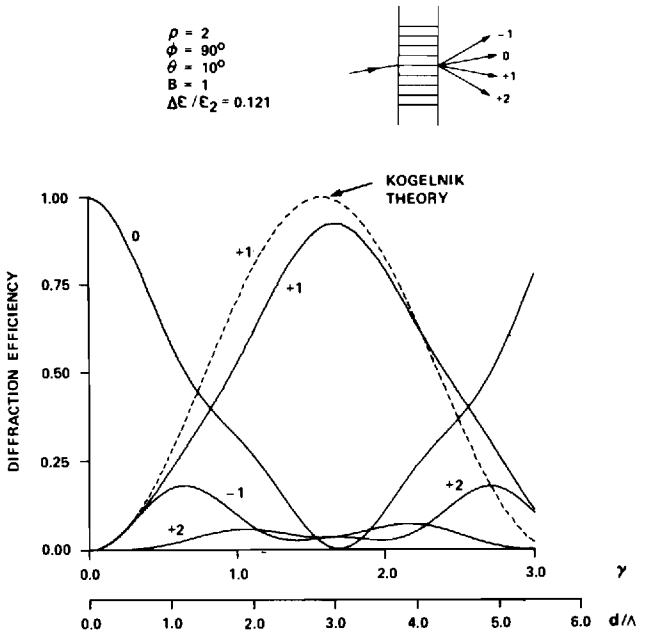


Fig. 2. The diffraction efficiencies of the transmitted waves for pure transmission grating (ϕ = 90°), with ρ = 2. The diffraction efficiencies of all reflected and transmitted waves not shown in the figure are less than 0.01. Here and in Figs. 3–6, $\epsilon_1 = \epsilon_2 = \epsilon_3$.

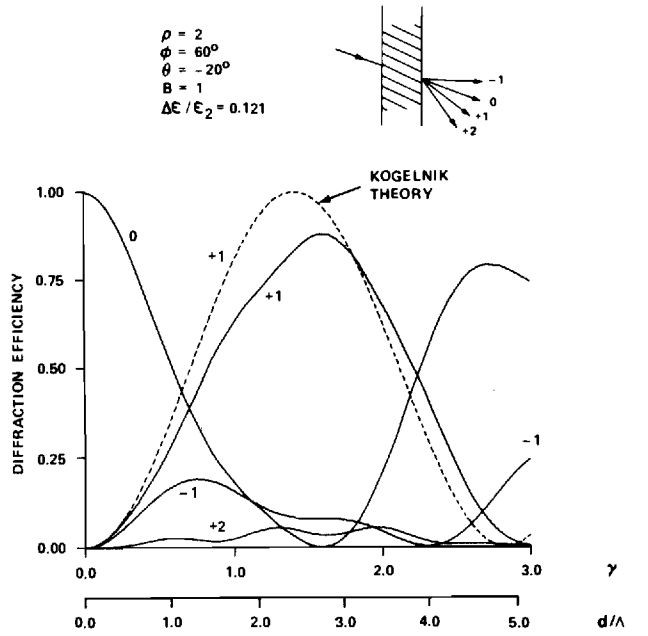


Fig. 3. The diffraction efficiencies of the transmitted waves for a ϕ = 60° slanted grating with ρ = 2. The diffraction efficiencies of all reflected and transmitted waves not shown in the figure are less than 0.01.

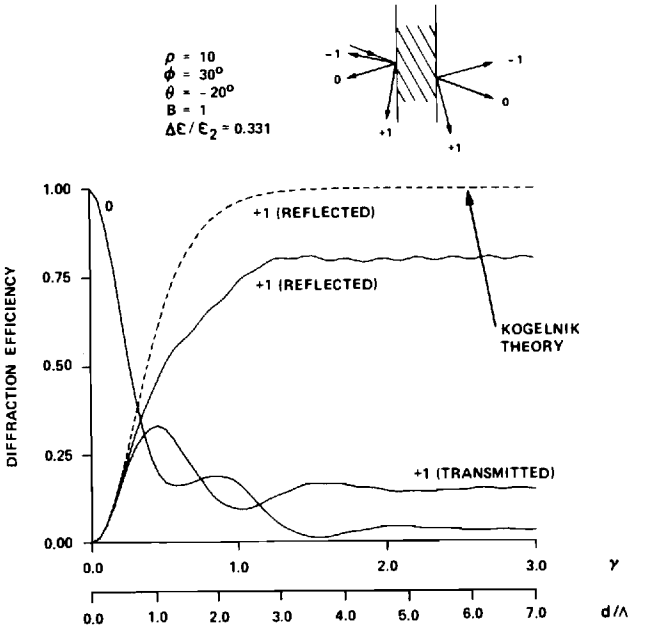


Fig. 4. The diffraction efficiencies for a ϕ = 30° grating with ρ = 10. The diffraction efficiencies of all reflected and transmitted waves not shown in the figure are less than 0.01.

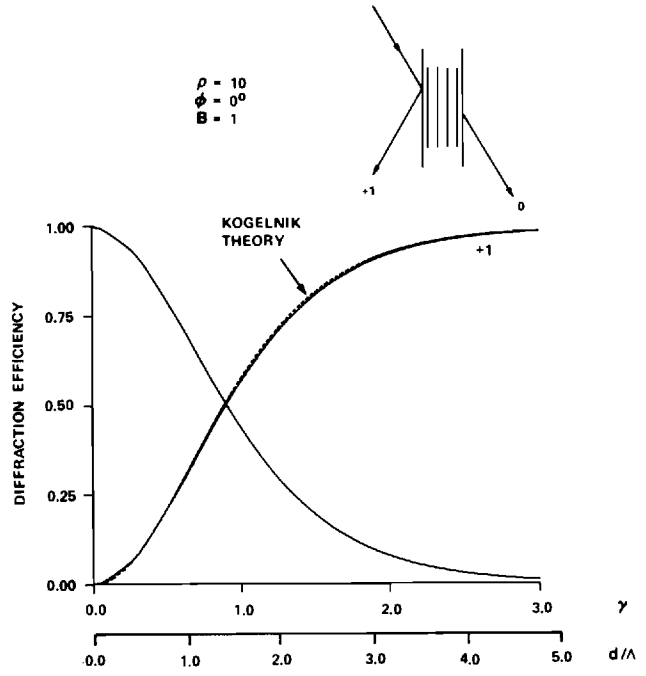


Fig. 5. The diffraction efficiencies of the transmitted and reflected waves at first-Bragg incidence (B = 1) for a pure reflection grating (ϕ = 0°) with ρ = 10. These two waves are the only waves that propagate in the unmodulated regions.

nar-grating diffraction problem by solving the wave equation in the three regions (Fig. 1) and then matching the tangential electric and the magnetic fields at the two boundaries to determine all the unknowns. The main difference between the two approaches is in the technique used to find solutions of the wave equation in the modulated region. In the present coupled-wave approach, the resulting system of coupled-wave equations is formulated into a simple matrix form for which the solution is readily obtained by calculating the eigenvalues and the eigenvectors of the coefficient matrix constructed from the coupled-wave equations. The eigenvalue problem, although not simple, has been extensively studied, and numerous efficient and straightforward computer programs to calculate the eigenvalues and the eigenvectors are available in typical computer-program libraries. The modal approach, on the other hand, requires that a transcendental relationship

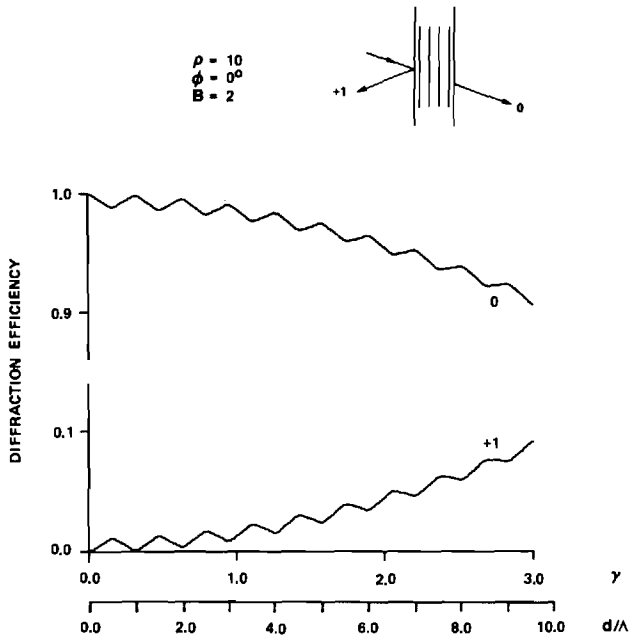


Fig. 6. The diffraction efficiencies of the transmitted and reflected waves at second-Bragg incidence ($B = 2$) for a pure reflection grating ($\phi = 0°$) with $\rho = 10$. These two waves are the only waves that propagate in the unmodulated regions.

in the form of a continued fraction expansion be solved to find the wavenumbers and their corresponding coefficients that are needed to solve the wave equation in the modulated region. This primary difficulty when applying the modal approach to the analysis of slanted gratings is due to the fact that there is no systematic technique yet available to solve this transcendental continued fraction relation for the general slanted-grating case.[8] For the modal approach the transcendental relationship can also be formulated as a matrix. This matrix is always $n \times n$ (as opposed to $2n \times 2n$ for the present coupled-wave analysis). Only for the unslanted, physically symmetric case is the resulting matrix for the modal approach in standard eigenvalue form. The corresponding wave numbers needed in the solution are the positive and negative square roots of the eigenvalues. The vector of coefficients, which is the same for both the positive and negative wave numbers, is the corresponding eigenvector of the problem. For slanted gratings the resulting modal-approach $n \times n$ matrix is not in the form of a standard eigenvalue problem. Therefore the wave numbers occurring in the matrix cannot be systematically determined.

B. Approximate Modal and Coupled-Wave Analyses

With the exception of the rigorous modal theory, previous analyses have typically been approximate. The present coupled-wave analysis is exact and rigorous, and it reduces to each of the previous analyses with the appropriate approximations and simplifications. For example, it reduces to Kong's coupled-wave analysis[6] and to Bergstein and Kermisch's analysis,[12] if a two-wave regime is assumed (that is, retaining only the first- and zero-order waves and neglecting all higher-order waves). Kogelnik's[4] analysis is obtained by assuming a two-wave regime, neglecting boundary diffraction, and neglecting second derivatives of the field amplitudes in Eq. (6). This last approximation, which is common in almost all previous coupled-wave analyses,[1-5,7] implies that the parameter $\Delta\epsilon/8\epsilon_2$ was assumed to be very small in these analyses. Magnusson and Gaylord's[7] analysis is obtained by neglecting boundary diffraction and the second derivatives of the field amplitudes. It is important to note that in previous coupled-wave analyses,[1-5,7] the amplitudes and intensities of the diffracted waves were calculated inside the modulated region. These analyses are based on solving the half-space grating problem. They do not (and cannot, because of the various approximations) solve the problem of general planar slab grating bounded by two different media.

C. Evaluation of Approximate Approaches

The rigorous coupled-wave theory presented in this paper is easily implemented and permits calculations to be performed rapidly. As such, this analysis is a powerful tool for analyzing the precise effect of using the various assumptions employed in the above approximate methods. It is difficult to quantify the error in general terms. However, the *trends* shown here have been consistently observed in the results of numerous calculations. Figures 7–10 show, for some typical cases, the absolute error in the primary +1 order that is due to using (a) the two-wave coupled-wave theory,[4] (b) the multiwave coupled-wave theory,[7] and (c) the two-wave modal theory.[12] In general, these grating structures do not behave simply as transmission gratings or reflection gratings. This is most obvious in slanted gratings (such as that shown in Fig. 4). A combination of transmission and reflection behaviors is typically present, even though one or the other dominates in each case, as determined by the location of the primary +1 diffracted order. For the slanted grating of Fig. 4, the +1 reflected order is the primary +1 diffracted order, and reflection behavior dominates in this case. Note that the +1 transmitted wave (which is phase matched to the +1 reflected wave) would have been assumed to be zero in the two-wave and multiwave (approximate) coupled-wave theories. This nonzero +1 transmitted diffracted wave is attributable entirely to the inclusion of second derivatives in this rigorous theory.

The gratings represented in Figs. 7–10 include the parameter values used in Figs. 2–5. The effect of neglecting boundary diffraction and second derivatives of field amplitudes is determined by the absolute error that results from using multiwave coupled-wave theory. As is shown in Figs. 7 and 8, this error is relatively small when forward-diffracted waves are dominant (transmission-grating behavior). However, when backward-diffracted waves are dominant (reflection-grating behavior), this error becomes relatively large, as is shown in Fig. 9. The effect of neglecting higher-order waves and thus retaining only one diffracted wave is determined by the absolute error that results in using two-wave modal theory. In a complementary way to the preceding results, Figs. 7 and 8 show that this error is large when forward-diffracted waves are dominant, and Fig. 9 shows that it is small when backward-diffracted waves are dominant. The results of many calculations support these conclusions. For example, calculations for two slanted gratings with exactly the same modulation, but one with transmission behavior dominant and the other with reflection behavior dominant, show that higher-order waves are more important in the transmission case and second derivatives and boundary diffraction are more important in the reflection case. Further, these basic conclu-

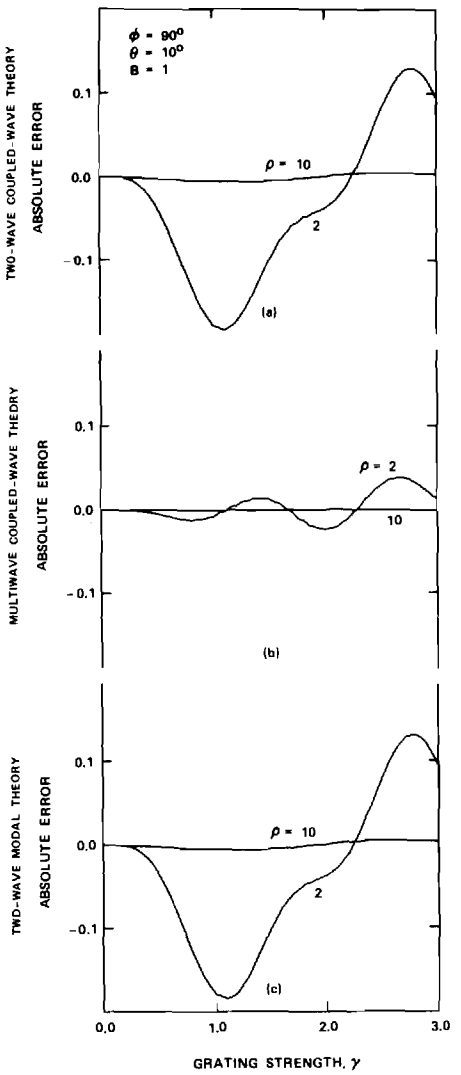


Fig. 7. Absolute error in the first-order diffraction efficiency as predicted by (a) two-wave coupled-wave theory, (b) multiwave coupled-wave theory, and (c) two-wave modal theory for the grating conditions shown in Fig. 2.

sions are supported by calculations for slanted gratings of equal modulation over a wide range of grating modulations. For example, at very small modulations, even though the errors that are due to the neglect of higher-order waves are small and the errors that are due to the neglect of second derivatives and boundary diffraction are small, the importance of these assumptions relative to each other is exactly the same as for large modulations. For equal grating modulations, higher-order waves are more important when transmission behavior dominates, and second derivatives and boundary diffraction are more important when reflection behavior dominates. This relative importance is unchanged regardless of whether the calculated errors are large or small. The combined effect of neglecting both higher-order waves and second derivatives and boundary diffraction is determined by the absolute error that results from using two-wave coupled-wave theory. This error is also shown in Figs. 7–9. From the two-wave coupled-wave theory error, it is consistently found that the total error that is due to neglecting both higher-order waves and second derivatives and boundary diffraction is approximately equal to the larger of the two constituent errors. The errors shown in Figs. 7–9 are all for the primary +1 order. However, numerous additional calculations for the other waves show a similar dependence on the approximations. For a large variety of cases analyzed, it was found in every case that if higher-order waves or boundary diffraction and second derivatives are important in accurately calculating the primary +1 order, then they are similarly important in accurately calculating the other orders.

For the case of a pure reflection grating ($\phi = 0$), the situation is somewhat more complicated. As ϕ approaches zero, all the forward-diffracted and all the backward-diffracted waves outside the modulated region converge to a single forward and a single backward direction, respectively. The directions of the various orders inside the modulated medium remain distinct. The error associated with the total backward-diffracted wave is shown in Fig. 10. In this case, the errors that are due to neglect of second derivatives and boundary diffraction are similar to those that are due to neglect of higher-order waves.

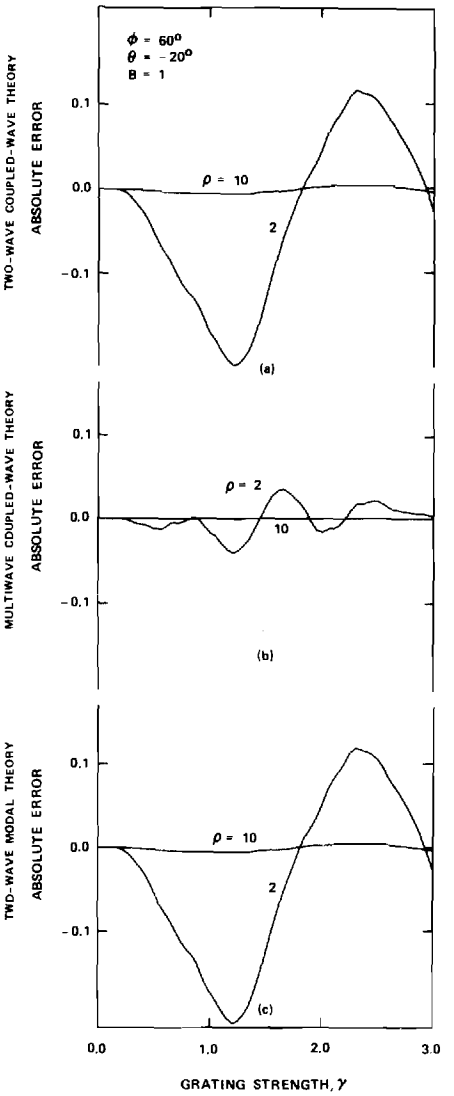


Fig. 8. Absolute error in the first-order diffraction efficiency as predicted by (a) two-wave coupled-wave theory, (b) multiwave coupled-wave theory, and (c) two-wave modal theory for the grating conditions shown in Fig. 3.

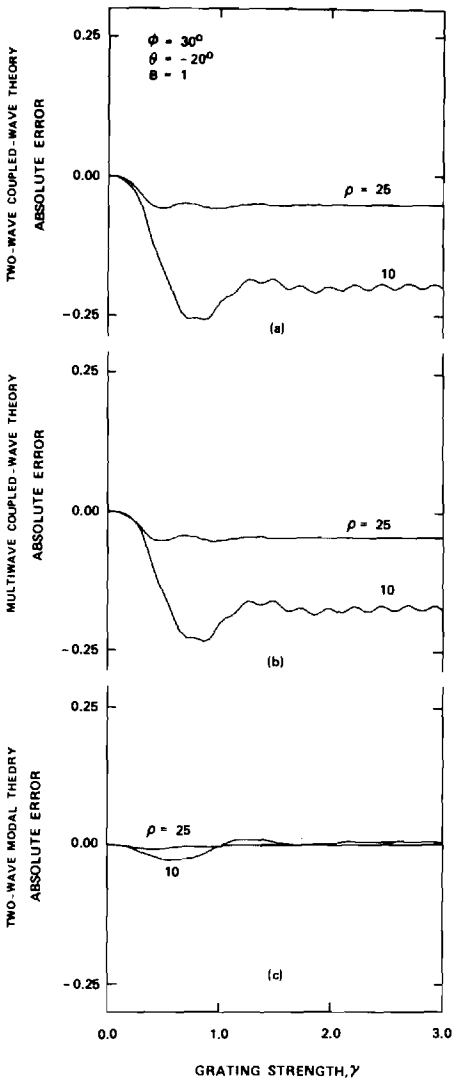


Fig. 9. Absolute error in the first-order diffraction efficiency as predicted by (a) two-wave coupled-wave theory, (b) multiwave coupled-wave theory, and (c) two-wave modal theory for the grating conditions shown in Fig. 4.

Incidence at the second-order Bragg condition is shown for a pure reflection grating in Fig. 6. It is important to note that solutions are not possible at even-order Bragg incidence ($B = 2, 4, 6, \ldots$) unless second derivatives and boundary diffraction are included in the analysis. For even-order Bragg conditions the coefficient of the first derivative term in Eq. (6) vanishes for pure reflection gratings. Thus calculations such as those shown in Fig. 6 are not even possible with most approximate theories.

5. DISCUSSION AND CONCLUSIONS

The diffraction of a plane electromagnetic wave incident obliquely upon a planar grating bounded by two different media has been analyzed by using an exact and rigorous coupled-wave approach. The grating vector may have any arbitrary angle with respect to the boundaries. The solution has been formulated in terms of state variables in a simple matrix form that is easily implemented on a digital computer. Sample rigorous calculations have been presented for transmission, reflection, and, for the first time, for general slanted gratings.

The present coupled-wave approach has been compared with the rigorous modal approach and with approximate modal and coupled-wave approaches. It was shown that inclusion of higher-order waves is more important than the inclusion of second derivatives and boundary diffraction for obtaining accurate predictions of diffraction efficiency when forward-diffracted waves are dominant (transmission-grating behavior). Conversely, it was shown that inclusion of second derivatives of the field amplitudes and boundary diffraction is more important for obtaining accurate predictions of diffraction efficiency when backward-diffracted waves are dominant (reflection-grating behavior).

The present method of analysis is useful in all applications in which planar gratings are utilized. However, it is especially valuable in applications in which slanted gratings and reflection gratings are used, such as in grating couplers and distributed-feedback lasers, since it is difficult to use the modal approach in these cases. This analysis may also be

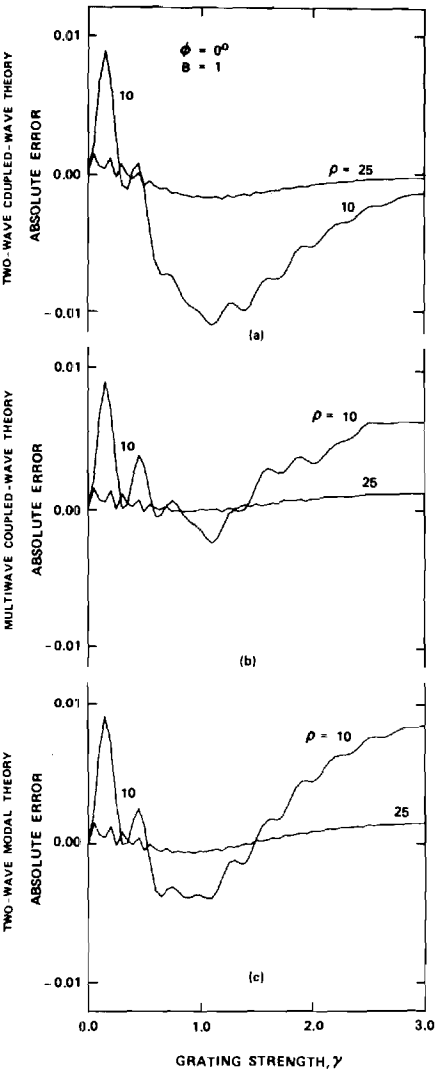


Fig. 10. Absolute error in the first-order diffraction efficiency as predicted by (a) two-wave coupled-wave theory, (b) multiwave coupled-wave theory, and (c) two-wave modal theory for the grating conditions shown in Fig. 5.

straightforwardly extended to absorption gratings and to mixed phase and absorption gratings, if desired.

APPENDIX: THE STATE EQUATION

The elements of the matrix $[b_{rs}]$ in Eq. (7), the state equation, are determined by Eq. (6), the governing set of differential equations. By defining the state variables as

$$S_{1,i} = S_i(u), \tag{A1}$$

$$S_{2,i} = \dot{S}_i(u), \tag{A2}$$

the constituent state equations may be written as

$$\dot{S}_{1,i} = S_{2,i}, \tag{A3}$$

$$\dot{S}_{2,i} = aS_{1,i+1} + b_iS_{1,i} + aS_{1,i-1} + c_iS_{2,i}, \tag{A4}$$

where $a = 8\epsilon_2/\Delta\epsilon$, $b_i = -a\rho i(i - B)$, and $c_i = a(\cos\theta' - i\mu\cos\phi)$. Writing Eqs. (A3) and (A4) as a single matrix state equation gives Eq. (7). In expanded form it is

$$\begin{bmatrix} \vdots \\ \dot{S}_{1,2} \\ \dot{S}_{1,1} \\ \dot{S}_{1,0} \\ \dot{S}_{1,-1} \\ \dot{S}_{1,-2} \\ \vdots \\ \dot{S}_{2,2} \\ \dot{S}_{2,1} \\ \dot{S}_{2,0} \\ \dot{S}_{2,-1} \\ \dot{S}_{2,-2} \\ \vdots \end{bmatrix} = \begin{bmatrix} & \vdots & & & & & \vdots & & & & \\ & 0\ 0\ 0\ 0\ 0 & & & 1\ 0\ 0\ 0\ 0 & & \\ & 0\ 0\ 0\ 0\ 0 & & & 0\ 1\ 0\ 0\ 0 & & \\ \cdots & 0\ 0\ 0\ 0\ 0 & \cdots & & 0\ 0\ 1\ 0\ 0 & \cdots & \\ & 0\ 0\ 0\ 0\ 0 & & & 0\ 0\ 0\ 1\ 0 & & \\ & 0\ 0\ 0\ 0\ 0 & & & 0\ 0\ 0\ 0\ 1 & & \\ & \vdots & & & \vdots & & \\ & b_2\ a\ 0\ 0\ 0 & & & c_2\ 0\ 0\ 0\ 0 & & \\ & a\ b_1\ a\ 0\ 0 & & & 0\ c_1\ 0\ 0\ 0 & & \\ \cdots & 0\ a\ b_0\ a\ 0 & \cdots & & 0\ 0\ c_0\ 0\ 0 & \cdots & \\ & 0\ 0\ a\ b_{-1}\ a & & & 0\ 0\ 0\ c_{-1}\ 0 & & \\ & 0\ 0\ 0\ a\ b_{-2} & & & 0\ 0\ 0\ 0\ c_{-2} & & \\ & \vdots & & & \vdots & & \end{bmatrix} \begin{bmatrix} \vdots \\ S_{1,2} \\ S_{1,1} \\ S_{1,0} \\ S_{1,-1} \\ S_{1,-2} \\ \vdots \\ S_{2,2} \\ S_{2,1} \\ S_{2,0} \\ S_{2,-1} \\ S_{2,-2} \\ \vdots \end{bmatrix}. \tag{A5}$$

This work was sponsored by the National Science Foundation under grant no. ECS-7919592 and by the Joint Services Electronics Program under grant no. DAAG29-78-C-0005.

REFERENCES

1. R. R. Aggrawal, "Diffraction of light by ultrasonic waves," Proc. Indian Acad. Sci. **31**, 417–426 (1950).
2. P. Phariseau, "On the diffraction of light by progressive supersonic waves," Proc. Indian Acad. Sci. Sect. A **44**, 165–170 (1965).
3. W. R. Klein and B. D. Cook, "Unified approach to ultrasonic light diffraction," IEEE Trans. Sonic Ultrason. **SU-14**, 123–134 (1967).
4. H. Kogelnik, "Coupled wave theory for thick hologram gratings," Bell Syst. Tech. J. **48**, 2909–2947 (1969).
5. G. L. Fillmore and R. F. Tynan, "Sensitometric characteristics of hardened dichromated-gelatin films," J. Opt. Soc. Am. **61**, 199–203 (1971).
6. J. A. Kong, "Second-order coupled-mode equations for spatially periodic media," J. Opt. Soc. Am. **67**, 825–829 (1977).
7. R. Magnusson and T. K. Gaylord, "Analysis of multiwave diffraction by thick gratings," J. Opt. Soc. Am. **67**, 1165–1170 (1977).
8. T. Tamir, H. C. Wang, and A. A. Oliner, "Wave propagation in sinusoidally stratified dielectric media," IEEE Trans. Microwave Theory Tech. **MTT-12**, 323–335 (1964).
9. T. Tamir and H. C. Wang, "Scattering of electromagnetic waves by a sinusoidally stratified half-space: I. Formal solution and analysis approximations," Can. J. Phys. **44**, 2073–2094 (1966).
10. T. Tamir, "Scattering of electromagnetic waves by a sinusoidally stratified half-space: II. Diffraction aspects of the Rayleigh and Bragg wavelengths," Can. J. Phys. **44**, 2461–2494 (1966).
11. C. B. Burckhardt, "Diffraction of a plane wave at a sinusoidally stratified dielectric grating," J. Opt. Soc. Am. **56**, 1502–1509 (1966).
12. L. Bergstein and D. Kermisch, "Image storage and reconstruction in volume holography," Proc. Symp. Modern Opt. **17**, 655–680 (1967).
13. R. S. Chu and T. Tamir, "Guided-wave theory of light diffraction by acoustic microwaves," IEEE Trans. Microwave Theory Tech. **MTT-18**, 486–504 (1970).
14. R. S. Chu and T. Tamir, "Wave propagation and dispersion in space–time periodic media," Proc. IEE **119**, 797–806 (1972).
15. F. G. Kaspar, "Diffraction by thick periodically stratified gratings with complex dielectric constant," J. Opt. Soc. Am. **63**, 37–45 (1973).
16. S. T. Peng, T. Tamir, and H. L. Bertoni, "Theory of periodic dielectric waveguides," IEEE Trans. Microwave Theory Tech. **MTT-23**, 123–133 (1975).
17. R. S. Chu and J. A. Kong, "Modal theory of spatially periodic media," IEEE Trans. Microwave Theory Tech. **MTT-25**, 18–24 (1977).
18. R. Magnusson and T. K. Gaylord, "Equivalence of multiwave coupled-wave theory and modal theory of periodic-media diffraction," J. Opt. Soc. Am. **68**, 1777–1779 (1978).
19. M. G. Moharam and T. K. Gaylord, "Coupled-wave analysis of reflection gratings," Appl. Opt., **20**, 240–244 (1981).
20. M. G. Moharam, T. K. Gaylord, and R. Magnusson, "Criteria for Bragg regime diffraction by phase gratings," Opt. Commun. **32**, 14–18 (1980).

Submitted
J. Opt. Soc. Am.

Rigorous Coupled-Wave Analysis of Grating Diffraction—

E Mode Polarization and Losses

M. G. Moharam and T. K. Gaylord
School of Electrical Engineering, Georgia Institute of Technology,
Atlanta, Georgia 30332

Received August , 1982

Rigorous coupled-wave theory of diffraction by dielectric gratings is extended to cover E mode polarization and losses. Unlike the H mode polarization case, it is shown that, in the E mode case, direct coupling exists between all diffracted orders rather than just between adjacent orders.

1. INTRODUCTION

Optical diffraction by dielectric gratings has been the subject of extensive, sustained research for many years. Fields of application include acousto-optics, integrated optics, quantum electronics, holography, and spectroscopy. Grating device functions include laser beam deflection, modulation, coupling, filtering, distributed feedback, distributed Bragg reflection, holographic beam combining, wavelength multiplexing, wavelength demultiplexing, and others.

A rigorous coupled-wave theory (without approximations) has recently been formulated for dielectric gratings.[1] This analysis applies for incident light of H mode polarization (electric field perpendicular to the plane of incidence and perpendicular to the grating vector). It is the purpose of this letter: (1) to show how the rigorous coupled-wave analysis can be extended to treat E mode polarization (electric field in the plane of incidence and in the plane of the grating vector) and lossy gratings, (2) to show that coupling exists between all diffracted orders for E mode polarization (unlike the case for H mode polarization in which the coupling is only between adjacent orders), and (3) to compare rigorous E mode results for gratings with and without losses to previous approximate E mode results, rigorous H mode results, and approximate H mode results.

2. GRATING WAVE EQUATIONS

A. General vector wave equations

The lossy dielectric grating is characterized by a relative permittivity that is periodic and is given by

$$\epsilon(x,z) = \hat{\epsilon}_o + \epsilon_1 \cos[K(x \sin\phi + z \cos\phi)] \tag{1}$$

where $\hat{\epsilon}_o$ is the average complex realtive permittivity given by

$$\hat{\epsilon}_o = \epsilon_o - j\sigma_o/\omega\epsilon_o \tag{2}$$

and ϵ_o is the average relative permittivity, σ_o is the average conductivity (representing the non-spatially-varying losses), ω is the optical radian frequency, ϵ_o is the permittivity of freespace, ϵ_1 is the amplitude of the sinusoidal relative permittivity, ϕ is the grating slant angle, K is the magnitude of the grating vector given by $K = 2\pi/\Lambda$, and Λ is the grating period. The planar boundaries of the grating are perpendicular to the z direction at z=0 and z=d. Although Eq. (1) represents the particular case of a sinusoidal permittivity, other grating profiles can also be treated. The electromagnetic fields inside a planar lossy dielectric grating with a spatially-varying relative permittivity are given by vector wave equations obtained directly from Maxwell's equations. The electric field vector wave equation is

$$\nabla^2\bar{E} + \nabla(\bar{E} \cdot \frac{\nabla\epsilon}{\epsilon}) + k^2\epsilon(x,z)\bar{E} = 0 \tag{3}$$

where $\bar{E}$ is the electric field, $\epsilon(x,z)$ is the periodic complex relative permittivity (dielectric constant), $k = 2\pi/\lambda$, and λ is the freespace wavelength. Similarly, the magnetic field vector wave equation is

$$\nabla^2\bar{H} + \frac{\nabla\epsilon}{\epsilon} \times \nabla \times \bar{H} + k^2\epsilon(x,z)\bar{H} = 0 \tag{4}$$

where $\overline{H}$ is the magnetic field. These general wave equations may be considerably simplified for particular incident wave polarizations.

B. H mode polarization wave equation

For H mode polarization (electric field perpendicular to plane of incidence and perpendicular to the grating vector), the electric field is solely in the y direction and so $\overline{E} = E\hat{y}$ where $\hat{y}$ is the unit vector in the y direction. Because the electric field is perpendicular to the grating modulation, then $\overline{E} \cdot \nabla\varepsilon = 0$. The electric field vector wave equation (3) therefore reduces to the scalar Helmholtz wave equation

$$\nabla^2 E + k^2 \varepsilon(x,z) E = 0 \tag{5}$$

This is the equation that is commonly solved in the analysis of dielectric grating diffraction.

C. E mode polarization wave equation

For E mode polarization, the electric field is in the plane of incidence and this plane contains the grating vector. The magnetic field is solely in the y direction and so $\overline{H} = H\hat{y}$. Because the magnetic field is only in the y direction, it is advantageous to select and to work with the magnetic field vector wave equation (4). This vector wave equation may be simplified using the vector identities $\nabla\varepsilon \times \nabla \times \overline{H} \equiv \nabla(\nabla\varepsilon \cdot \overline{H}) - (\nabla\varepsilon \cdot \nabla)\overline{H} - (\overline{H} \cdot \nabla)\nabla\varepsilon - \overline{H} \times (\nabla \times \nabla\varepsilon)$ and $\nabla \times \nabla\varepsilon \equiv 0$. For E mode polarization, H is perpendicular to $\nabla\varepsilon$ and thus $\nabla\varepsilon \cdot \overline{H} = 0$ and $(\overline{H} \cdot \nabla)\nabla\varepsilon = 0$. The magnetic field vector wave equation thus reduces to

$$\nabla^2 H - \left(\frac{\nabla\varepsilon}{\varepsilon} \cdot \nabla\right) H + k^2 \varepsilon(x,z) H = 0 \tag{6}$$

This equation contains an additional term in comparison to the E mode wave equation (5).

3. COUPLED-WAVE EQUATIONS

A. H mode coupled-wave equations

The H mode polarization coupled-wave equations may be obtained by expanding the electric field in space harmonics as[1]

$$E(x,z) = \sum_{i=-\infty}^{+\infty} S_i(z) \exp(-j\bar{\sigma}_i \cdot \bar{r}) \tag{7}$$

where i is the integer space harmonic index, $S_i(z)$ is the space harmonic electric field amplitude, $\bar{\sigma}_i = \bar{k}_2 - i\bar{K}$ from the Floquet theorem, and $\bar{k}_2$ is the wavevector of the zero-order (i=0) refracted wave in region 2, the grating region $0 \leq z \leq d$. (Region 1 is the input region $z \leq 0$ and region 3 is the output region $z \geq d$.) The magnitude of $\bar{k}_2$ is $k_2 = 2\pi(\hat{\varepsilon}_o)^{1/2}/\lambda$. Each space harmonic $S_i(z)$ inside the grating is phase matched to a forward-diffracted and a backward-diffracted wave. These waves may be either propagating or evanescent. Substitution of (7) into (5) leads to an infinite exponential series in terms of $S_i(z)$. Each coefficient may be expressed as a function of i and z and each exponent as a function of i and x. For nontrivial solutions, each coefficient must be equal to zero. This gives the coupled-wave equations. After simplification the H mode coupled-wave equations are

$$\frac{1}{2\pi^2} \frac{d^2S_i(z)}{dz^2} - j\frac{2}{\pi}\left[\frac{(\hat{\varepsilon}_o - \varepsilon_I \sin\theta')^{1/2}}{\lambda} - \frac{i\cos\phi}{\Lambda}\right]\frac{dS_i(z)}{dz} +$$

$$+ \frac{2i(m-i)}{\Lambda^2} S_i(z) + \frac{\varepsilon_1}{\lambda^2}[S_{i+1}(z) + S_{i-1}(z)] = 0 \tag{8}$$

where θ' is the angle of incidence in region 1 of the input plane wave, ε_I is the average relative permittivity in region 1, and m is defined as

$$m \equiv 2(\Lambda/\lambda)[\varepsilon_I^{\frac{1}{2}} \sin\phi \sin\theta' + (\hat{\varepsilon}_o - \varepsilon_I \sin^2\theta')^{\frac{1}{2}} \cos\phi] \ . \qquad (9)$$

When the real part of m is an integer, this represents a Bragg condition. These rigorous coupled-wave equations may be solved by the state variable methods[2] and together with the appropriate boundary conditions, all of the diffracted fields may be determined.[1]

B. E mode coupled-wave equations

The vectorial E mode wave equation (6) can also be reformulated as a set of scalar coupled-wave equations. The vector term may be expanded as

$$-\left(\frac{\nabla\varepsilon}{\varepsilon}\cdot\nabla\right)H = \frac{\varepsilon_1 \sin(\bar{K}\cdot\bar{r})}{\varepsilon_o + \varepsilon_1 \cos(\bar{K}\cdot\bar{r})}\left(\sin\phi\,\frac{\partial H}{\partial x} + \cos\phi\,\frac{\partial H}{\partial z}\right)\frac{2\pi}{\Lambda} \qquad (10)$$

and thus only a y component equation exists. To put this term into standard form, the leading factor is expanded in a Fourier series as

$$\frac{\varepsilon_1 \sin(\bar{K}\cdot\bar{r})}{\varepsilon_o + \varepsilon_1 \cos(\bar{K}\cdot\bar{r})} = -j \sum_{h=-\infty}^{+\infty} A_h \exp(jh\,\bar{K}\cdot\bar{r}) \qquad (11)$$

where $A_h = -\{[(\hat{\varepsilon}_o/\varepsilon_1)^2 - 1]^{\frac{1}{2}} - (\hat{\varepsilon}_o/\varepsilon_1)\}^h$ for $h \geq 1$, $A_{-h} = -A_h$, and $A_o = 0$. For the E mode case the magnetic field is expanded in space harmonics as

$$H(x,z) = \sum_{i=-\infty}^{+\infty} U_i(z) \exp(-j\bar{\sigma}_i\cdot\bar{r}) \qquad (12)$$

where $U_i(z)$ is the space harmonic magnetic field amplitude and the other

quantities are defined as before. Substituting (10), (11), and (12) into (6) and proceeding as before gives the E mode coupled-wave equations as

$$\frac{1}{2\pi^2}\frac{d^2U_i(z)}{dz^2} - j\frac{2}{\pi}\left[\frac{(\hat{\varepsilon}_o - \varepsilon_I \sin\theta')^{\frac{1}{2}}}{\lambda} - \frac{i\cos\phi}{\Lambda}\right]\frac{dU_i(z)}{dz}$$

$$- j\frac{\cos\phi}{\pi\Lambda}\sum_h A_h \frac{dU_{i-h}(z)}{dz} + \frac{2i(m-i)}{\Lambda^2}U_i(z) + \frac{\varepsilon_1}{\lambda^2}[U_{i+1}(z) + U_{i-1}(z)]$$

$$+ \frac{2}{\Lambda^2}\sum_h (i - h - \frac{m}{2}) A_h U_{i-h}(z) = 0 \qquad (13)$$

These equations for E mode polarization are clearly more complicated than the H mode coupled-wave equations (8). The two additional terms in (13) both contain a series in A_h and U_{i-h}. Whereas the H mode equations only contain S_{i-1}, S_i, and S_{i+1} amplitude terms, the E mode equations contain U_{i-1}, U_i, U_{i+1}, and U_{i-h} amplitude terms. Therefore in the H mode case there is direct coupling only between adjacent orders, but in the E mode case there is direct coupling between all of the diffracted orders. Although the number of terms in the E mode case is larger, the resulting coupled-wave equations may be solved by the state variables method in exactly the same manner as in the H mode case.

4. BOUNDARY CONDITIONS

At the boundaries of the grating (z=0 and z=d), the tangential components of the electric field and magnetic field must be continuous. In this way the field of each diffracted-order outside of the grating volume is related to the corresponding space harmonic field inside the grating. Thus in order to construct the boundary conditions, the tangential components of E and H must be determined.

A. H mode polarization tangential fields

For H mode polarization, the tangential component of the electric field is the y component of E and it is given by Eq. (7) directly. The values of $S_i(0)$ and $S_i(d)$ needed in Eq. (7) are obtained by solving the H-mode coupled-wave equations (8) for $S_i(z)$. The tangential component of the magnetic field is the x component of H. It may be obtained from the Maxwell curl equation $\nabla \times \bar{E} = -\partial\bar{B}/\partial t$. The result is $H_x = (-j/\omega\mu)\ \partial E_y/\partial z$ and together with Eq. (7) the tangential magnetic field is

$$H_x = (-j/\omega\mu)\ \frac{\partial}{\partial z} \sum_{i=-\infty}^{+\infty} S_i(z)\ \exp(-j\bar{\sigma}_i \cdot \bar{r})\ . \tag{14}$$

B. E mode polarization tangential fields

For E mode polarization, the tangential component of the magnetic field is the y component of H and it is given by Eq. (12) directly. The values of $U_i(0)$ and $U_i(d)$ to be used in Eq. (12) are obtained by solving the E mode coupled-wave equations (13) for $U_i(z)$. The tangential electric field is the x component of E. It may be obtained from the other Maxwell curl equation $\nabla \times \bar{H} = \partial\bar{D}/\partial t$. The result is $E_x = [j/\omega\epsilon_o\ \epsilon(x,z)]\ \partial H_y/\partial z$. Expanding $1/\epsilon(x,z)$ into a Fourier series gives

$$\frac{1}{\epsilon(x,z)} = \sum_{h=-\infty}^{+\infty} G_h\ \exp(jh\ \bar{K} \cdot \bar{r}) \tag{15}$$

where $G_h = \{[(\hat{\epsilon}_o/\epsilon_1)^2 - 1]^{1/2} - (\hat{\epsilon}_o/\epsilon_1)\}^{|h|}/(\hat{\epsilon}_o^2 - \epsilon_1^2)^{1/2}$. Substituting this into the above equation for E_x and using H_y as given by Eq. (12), yields the tangential electric field as

$$E_x = (j/\omega\epsilon_o) \sum_{i=-\infty}^{+\infty} \exp(-j\bar{\sigma}_i \cdot \bar{r}) \sum_{h=-\infty}^{+\infty} G_h \left[\frac{dU_{i-h}(z)}{dz} - j(\bar{\sigma}_{i-h} \cdot \hat{z})U_{i-h}(z)\right] \tag{16}$$

5. DISCUSSION

The rigorous scalar coupled-wave equations describing diffraction by planar lossy dielectric gratings have been presented for both H mode and E mode polarizations as derived from the general vector wave equations. The resulting coupled-wave equations for both cases can be solved in the same manner using state variable methods.[2] Using these in combination, any arbitrary input polarization may thus be treated.

For H mode polarization, the well-known result of direct coupling only between adjacent diffracted orders is obtained. However, for E mode polarization it has been shown that direct coupling exists between all diffracted orders. The set of boundary condition equations for each polarization is a set of linear algebraic equations and after the coupled-wave equations are solved, these may then be solved for the phase-matched propagating and evanescent wave amplitudes outside of the grating.

Numerous calculations have been performed to obtain the fundamental and higher-order forward and backward diffracted wave amplitudes for both H mode polarization and E mode polarization incident waves. Results for an example lossless situation are shown in Fig. 1. The rigorously-calculated diffraction efficiencies of several forward-diffracted orders are shown in Fig. 1(a). The power in the E mode diffracted waves is initially less than that for the H mode polarization due to the reduced coupling in E mode compared to H mode. However as the thickness is increased the E mode fundamental (+1) diffraction efficiency exceeds the corresponding H mode diffraction efficiency. Diffraction efficiency results from Kogelnik's two-wave first-order coupled-wave theory[3] are shown for comparison. In this theory, the higher-order waves are neglected and so only the transmitted (0) and fundamental (+1) orders can be shown in Fig. 1(b). The

rigorously-calculated diffraction efficiencies are seen to be generally less than those predicted by two-wave first-order coupled-wave theory. In addition, the E mode fundamental diffraction efficiency has been shown to have a larger maximum value than H mode in this case.

The diffraction efficiencies for the same grating but with the addition of losses are shown in Fig. 2. The non-zero conductivity produces an average absorption that is seen to cause a reduction in all of the diffracted intensities as would be anticipated.

In separate calculations, it was found that both H mode and E mode diffraction efficiencies reduce to the values predicted by Kogelnik's two-wave first-order coupled-wave theory in the limit of small modulation. In other rigorously-calculated E mode results, it was verified that the first-order diffraction efficiency goes to zero when the zero-order space harmonic field in the grating is perpendicular to the fundamental first-order (+1) space harmonic field. For E mode polarization this means that the electric field vectors inside the grating associated with the transmitted and diffracted waves are perpendicular to each other and thus there is no coupling between them. Therefore the diffraction efficiency is zero. These characteristics are, of course, expected and serve as additional verification of the analysis.

The generalized rigorous coupled-wave analysis presented here is mathematically exact. There are no theoretical deficiencies or approximations. Any arbitrary level of accuracy is obtainable by increasing the number of orders retained in the analysis. However, convergence is very rapid. In the numerical calculations presented here, the diffracted amplitudes were determined to one part in 10^8. Conservation of power among the beams was accurate to one part in 10^{12}. It should be recognized that

this level of accuracy greatly exceeds that usually presented in grating diffraction calculations.

This work was supported by the National Science Foundation and by the Joint Services Electronics Program.

REFERENCES

1. M. G. Moharam and T. K. Gaylord, "Rigorous coupled-wave analysis of planar-grating diffraction," J. Opt. Soc. Am. 71 811-818 (1981).

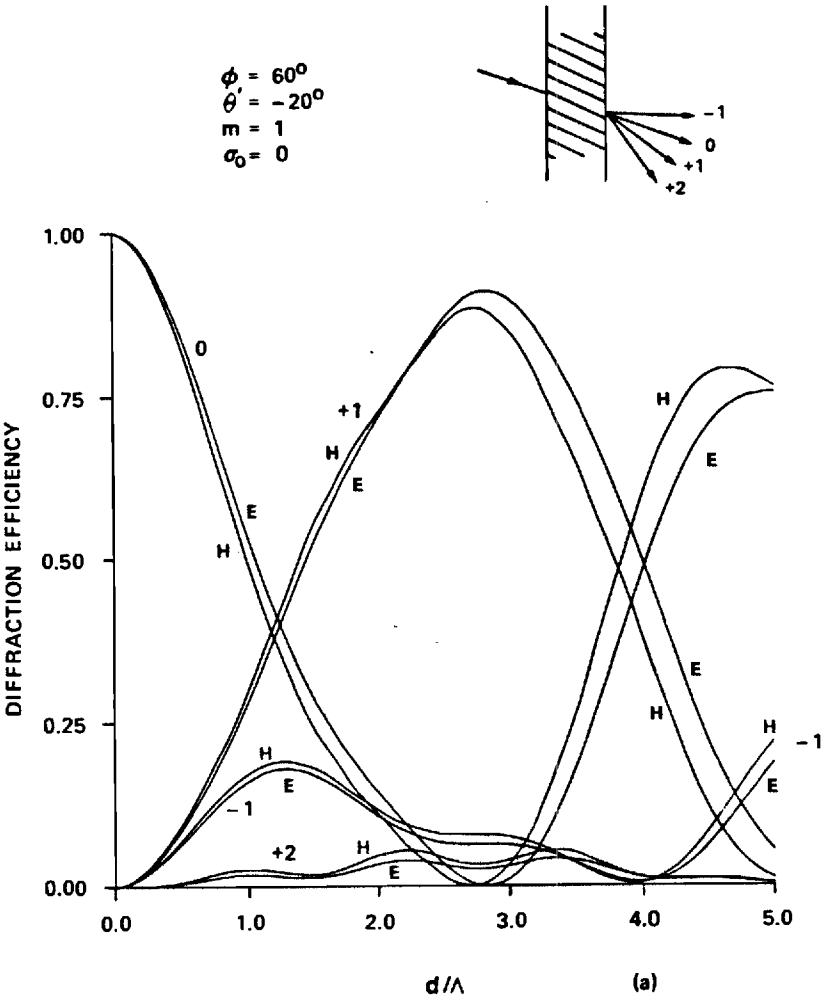
2. e.g. C. L. Liu and J. W. S. Liu, Linear systems analysis (McGraw-Hill, New York, 1975).

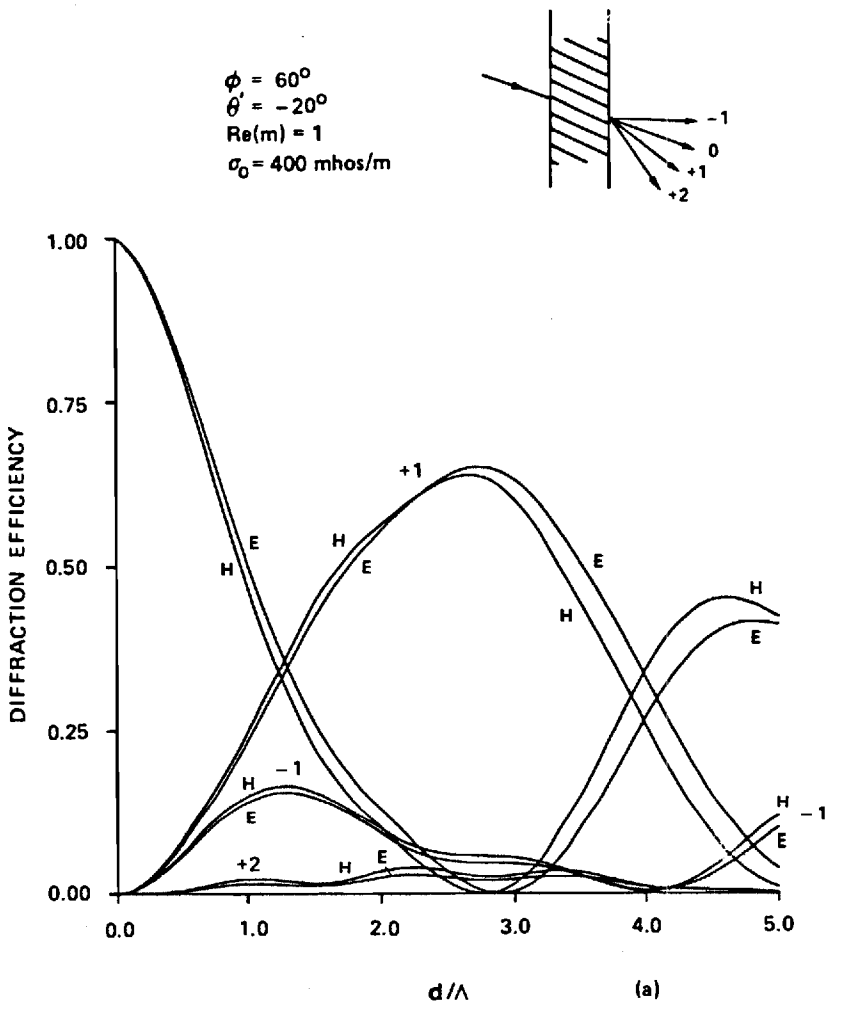
3. H. Kogelnik, "Coupled wave theory for thick hologram gratings," Bell Syst. Tech J. 48 2909-2947 (1969).

FIGURE CAPTION

Fig. 1 Diffraction efficiencies of forward-diffracted waves for a lossless $\phi = 60°$ slanted grating. The diffraction efficiencies for all diffracted waves not shown are less than 0.01. The average permittivity inside and outside the grating is the same. The angle of incidence $\theta' = -20°$ is at the first Bragg angle (m=1). The modulation is $\varepsilon_1/\varepsilon_o = 0.120$. (a) Rigorously calculated results for both H mode and E mode polarizations. (b) Two-wave first-order coupled-wave theory results for both H mode and E mode polarizations.

Fig. 2 Diffraction efficiencies of forward-diffracted waves for the same grating as Fig. 1 but with losses present. The average conductivity is $\sigma_o = 400\ (\text{ohm-m})^{-1}$. Other parameters are $\lambda = 514.5$ nm, $\Lambda = 0.9876$ microns, and $\varepsilon_o = 2.25$. The average permittivity outside of the grating is the same as that inside. (a) Rigorously calculated results for both H mode and E mode polarizations. (b) Two-wave first-order coupled-wave theory results for both H mode and E mode polarizations.


$\phi = 60^o$
$\theta' = -20^o$
m = 1
$\sigma_0 = 0$
-1
0
+1
+2
DIFFRACTION EFFICIENCY
1.00
0.75
0.50
0.25
0.00
0.0
1.0
2.0
3.0
4.0
5.0
d/Λ
(a)
(b)


ϕ = 60°
θ' = −20°
Re(m) = 1
σ_0 = 400 mhos/m
−1
0
+1
+2
DIFFRACTION EFFICIENCY
1.00
0.75
0.50
0.25
0.00
0.0
1.0
2.0
3.0
4.0
5.0
d/Λ
(a)
DIFFRACTION EFFICIENCY
1.00
0.75
0.50
0.25
0.00
0.0
1.0
2.0
3.0
4.0
5.0
d/Λ
(b)

Reprinted from **Applied Optics,** Vol. 20, page 240, January 15, 1981

Coupled-wave analysis of reflection gratings

M. G. Moharam and T. K. Gaylord

Diffraction of electromagnetic waves by longitudinally periodic media (reflection gratings) is rigorously analyzed using an exact coupled-wave approach. The analysis is formulated in a simple matrix form easily implemented on a digital computer. The intensities of the two waves, the diffracted (reflected) and transmitted waves, are calculated for a wide range of parameters. These exact results are then compared to results obtained with approximate two-wave modal, and with multiwave and two-wave coupled-wave analyses. The applicability of these approximate theories is discussed. The case of other than first-order Bragg incidence is also included.

I. Introduction

Planar gratings have applications in many different areas such as acoustooptics, integrated optics, holography, and spectroscopy. Consequently, the diffraction of electromagnetic waves by these periodic structures has been extensively studied in recent years.

Most of the previous work, whether utilizing the rigorous modal approach[1–10] or the simpler but approximate coupled-wave approach,[11–15] have been for transversely periodic media. These are unslanted transmission gratings (the direction of the periodicity is parallel to the grating surface). However, recent applications such as grating couplers and distributed feedback lasers are either slanted or reflection type gratings. Kogelnik[13] and Magnusson and Gaylord[14] have analyzed general slanted gratings using the two-wave and the multiwave coupled-wave analyses, respectively. Their analyses are not exact because of approximations introduced into the coupled-wave approach such as (1) neglecting higher-order waves,[13] (2) neglecting the second derivatives of the field amplitudes,[13,14] and (3) neglecting the boundary diffracted waves.[13,14] Bergstein and Kermisch[5] analyzed slanted gratings using the modal approach, but their analysis is for very small modulations (two-wave approximation). Chu and Kong[9] have studied the general slanted grating problem using an exact modal approach, but they did not present calculated results for reflection or slanted gratings because of the extremely difficult numerical problem resulting from the modal formulation.[16]

In this paper, reflection gratings bounded by two different media are analyzed rigorously using an exact coupled-wave approach that has recently been developed.[16] The analysis is formulated in a simple matrix form and numerical solutions are easily obtained. The intensities of the diffracted and the transmitted waves are calculated for a wide range of parameters. These exact results are compared with those from the approximate two-wave modal[5] and two-wave and multiwave coupled-wave analyses of Kogelnik[13] and of Magnusson and Gaylord.[14] The applicability of these approximate theories is discussed. The case of light incidence at other than first-order Bragg condition is also included.

II. Theory and Method of Solution

The grating under consideration is a lossless (pure phase) sinusoidal permittivity grating with an obliquely incident wave polarized perpendicular to the plane of incidence (H-mode). The modulated region is bounded by two different media with relative permittivities ϵ_1 and ϵ_3, as shown in Fig. 1. All three regions have the permeability of free space. The relative permittivity in the modulation region of thickness d is given by

$$\epsilon(x,z) = \epsilon_2 + \Delta\epsilon \cos[K(x \sin\phi - z \cos\phi)], \quad (1)$$

where $K = 2\pi/\Lambda$, and Λ is the grating period. The quantities ϵ_2 and $\Delta\epsilon$ are the average relative permittivity and amplitude of the modulation in the relative permittivity. The slant angle is ϕ, and for the case of pure reflection gratings, ϕ is zero. In the modal and coupled-wave analyses, ϕ approaches zero but is not equal to zero because that causes the grating not to be

The authors are with Georgia Institute of Technology, School of Electrical Engineering, Atlanta, Georgia 30332.

Received 5 July 1980.
0003-6935/81/020240-05$00.50/0.

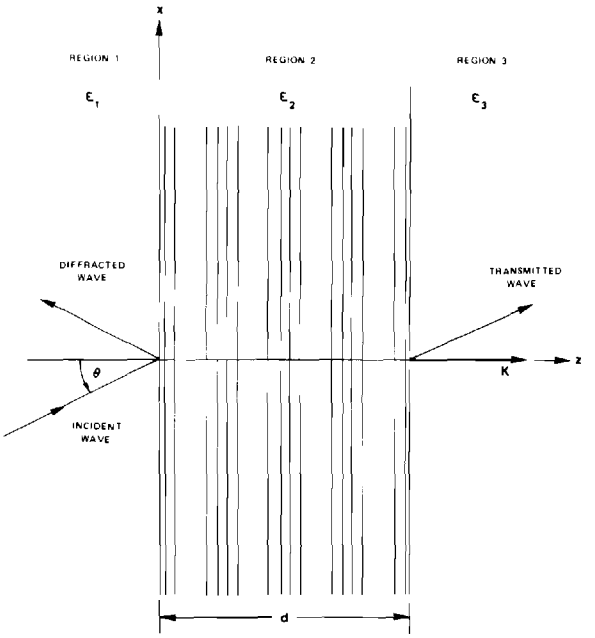


Fig. 1. Geometry for diffraction of an electromagnetic wave by a planar reflection grating.

strictly periodic and the Floquet theorem, which is the basis for the modal and the coupled-wave analyses, will not apply. As outlined in Ref. 16 the assumed solutions of the wave equation in the two unmodulated and the modulated regions for the normalized electric field are:

In region 1 ($z < 0$),

$$E_1 = \exp[-j(\beta_0 x + \xi_{10} z)] + \sum_i R_i \exp[-j(\beta_i x - \xi_{1i} z)]. \quad (2)$$

In region 3 ($z > d$),

$$E_3 = \sum_i T_i \exp\{-j[\beta_i x + \xi_{3i}(z - d)]\}. \quad (3)$$

In region 2 ($0 < z < d$),

$$E_2 = \sum_i \hat{S}_i(z) \exp[-j(\beta_i x + \xi_{2i} z)], \quad (4)$$

where $\beta_i = k_1 \sin\theta - iK \sin\phi$; $\xi_{li} = (k_l^2 - \beta_i^2)^{1/2}$ for $l = 1, 3$; $\xi_{2i} = k_2 \cos\theta' - iK \cos\phi$; $k_l = 2\pi(\epsilon_l)^{1/2}/\lambda$ for $l = 1, 2, 3$, where λ is the free-space wavelength of the light, θ is the angle of incidence in region 1, and θ' is the angle of refraction in region 2. R_i and T_i are the amplitudes of ith reflected and ith transmitted waves and are to be determined. $\hat{S}_i(z)$ is the amplitude of the ith wave anywhere in the modulated region and is to be determined by solving the wave equation

$$\nabla^2 E_2 + (2\pi/\lambda)^2 \epsilon(x,z) E_2 = 0, \quad (5)$$

in the modulated region. It is important to note that by taking ϕ approaching, but not equal to, zero all of the individually calculated values of β_i converge to the same value. That is, all of the reflected waves (R_i) and all of the transmitted waves (T_i) have virtually the same propagation constant but are still individually distinguishable. With an understanding of this limiting process, ϕ can be taken to be equal to zero in the calculations.

To find $\hat{S}_i(z)$, Eqs. (1) and (4) are substituted into Eq. (5) with ($\phi = 0$) resulting in the system of coupled-wave equations with $S_i(v) = \hat{S}_i(z)$:

$$\frac{1}{\rho B^2}\frac{d^2 S_i(v)}{dv^2} = [1 - (2i/B)]\frac{dS_i(v)}{dv} - \rho i(i - B)S_i(v) + S_{i+1}(v) + S_{i-1}(v), \quad (6)$$

where $\rho = 2\lambda^2/(\wedge^2 \Delta\epsilon)$, $B = 2 \wedge \epsilon_2^{1/2} \cos\theta'/\lambda$, and $v = j\pi\Delta\epsilon z/(2\lambda\epsilon_2^{1/2}\cos\theta') = j\gamma(z/d)$. Here B is the Bragg condition parameter, which is unity for first Bragg incidence. The quantity ρ is the regime parameter, which is shown in this paper to determine the boundary for the two-wave Bragg diffraction regime ($\rho > 10$) for reflection gratings. Thus, this parameter is shown to have a significance for reflection gratings similar to its significance in transmission gratings.[17]

Equation (6) can be written in matrix "state equation" form[16] as

$$\begin{bmatrix} \mathbf{S}' \\ \mathbf{S}'' \end{bmatrix} = [b_{rs}] \begin{bmatrix} \mathbf{S} \\ \mathbf{S}' \end{bmatrix}, \quad (7)$$

where $\mathbf{S}$, $\mathbf{S}'$, and $\mathbf{S}''$ are the column vectors $S_i(v)$, $dS_i(v)/dv$, and $d^2S_i(v)/dv^2$, respectively, and $[b]$ is the coefficient matrix determined from Eq. (6). The solution of Eq. (7) gives the field amplitude $S_i(v)$ as

$$S_i(v) = \sum_m C_m w_{im} \exp(q_m v), \quad (8)$$

where q_m are the eigenvalues of the matrix $[w]$ composed by the eigenvectors of the coefficient matrix $[b]$. Note that w_{im} is the mth element of the row of the matrix $[w]$ that corresponds to the ith wave (not the ith row). C_m together with T_i and R_i are determined by matching the tangential electric and magnetic fields at the two boundaries. The boundary conditions give[16]

$$2\delta_{i0}\xi_{1i}d = \sum_m C_m w_{im}(\xi_{2i}d + \xi_{1i}d - q_m\gamma), \quad (9)$$

$$0 = \sum_m C_m w_{im}(\xi_{2i}d - \xi_{3i}d - q_m\gamma)\exp(jq_m\gamma), \quad (10)$$

where $\xi_{li}d = \rho\gamma B^2(\epsilon_l/\epsilon_2 - \sin^2\theta')^{1/2}/2\cos\theta'$ for $l = 1,3$, and $\xi_{2i}d = \rho\gamma B(B - 2i)/2$. The values of C_m are found by solving this system of linear equations. Then R_i and T_i are calculated from

$$\delta_{i0} + R_i = \sum_m C_m w_{im}, \quad (11)$$

$$T_i = \sum_m C_m w_{im} \exp[j(q_m\gamma - \xi_{2i}d)], \quad (12)$$

where δ_{i0} is the Kronecker delta function. As ϕ approaches zero all the reflected waves collapse into one wave of intensity $\Sigma_i R_i R_i^*$ and all the transmitted waves collapse into one wave of intensity $\Sigma_i T_i T_i^*$. Clearly the sum of these two individual sums should be equal to the input intensity [normalized to unity in Eq. (2)]. Therefore, there may be many waves in the modulated

region, but there is only one reflected and one transmitted wave in the two unmodulated regions as ϕ approaches zero.

III. Comparison to Previous Coupled-Wave Analyses

Although many authors state that their analyses apply to reflection gratings, very few results have been presented for these gratings. Authors[9] using the modal approach to analyze reflection gratings did not present any calculated results because solutions of the modal formulation are extremely difficult to obtain for this case. In most coupled-wave analyses, the second derivatives of the field amplitudes and the boundary diffraction are neglected. That is, coupled-wave analyses calculate the wave amplitudes inside the grating, not those of the diffracted waves leaving the grating. In the present analysis both the wave amplitudes inside and outside the grating are calculated. Kogelnik[13] analyzed reflection gratings using the coupled-wave approach where he assumed a two-wave regime in addition to the other approximations mentioned above. Our present analysis reduces to his analysis in the limit when $1/\rho B^2 = 0$ and retaining only the zero-order and first-order waves in Eq. (6). Kogelnik[13] obtained closed form expressions for the field amplitudes in the modulated region, which may be written in terms of our parameters as

$$S_0(v) = \cosh(\gamma + jv)/\cosh\gamma, \tag{13}$$

$$S_1(v) = -j \sinh(\gamma + jv)/\cosh\gamma, \tag{14}$$

$$S_i(v) = 0, \quad i \neq 0, 1. \tag{15}$$

In this case, the reflected intensity

$$I_r = R_1 R_1^* = S_1(0) S_1^*(0) = \tanh^2\gamma. \tag{16}$$

The transmitted intensity is the difference between the incident intensity and the reflected intensity. Note, that in obtaining S_0 and S_1 Kogelnik assumed that $S_0(0) = 1$ and $S_1(j\gamma) = 0$ (thus, neglecting boundary reflections). Magnusson and Gaylord[14] extended Kogelnik's analysis to allow for all the higher-order waves in the modulated region. They neglected boundary diffraction and the second derivatives of the field amplitudes. The present analysis reduces to their analysis if $1/\rho B^2 = 0$ in Eq. (6). They also did not present calculated results for reflection gratings. But, if one calculates all the S_i field amplitudes in the modulated region, then the intensity of the reflected wave leaving the grating is

$$I_r = \sum_{i>0} (2i - 1) S_i(0) S_i^*(0). \tag{17}$$

Note that to obtain S_i Magnusson and Gaylord[14] assumed $S_i(j\gamma) = 0$ for $i > 0$, $S_i(0) = 0$ for $i < 0$, and $S_0(0) = 1$, thus neglecting boundary diffraction. It is important to note that in these approximate coupled-wave analyses that neglect the second derivatives of the field amplitudes, solutions are not possible for incidence at even-order Bragg conditions (i.e., $B = 2, 4, 6, \ldots$) because for even-order Bragg conditions, the first derivative of ith wave also vanishes [$1 - 2i/B = 0$ in Eq. (6)]. The results of the Bergstein and Kermisch[5] two-wave modal analysis can be obtained from the present formulation by keeping only the zero-order and first-order waves in Eqs. (2) through (4) and neglecting all higher-order waves.

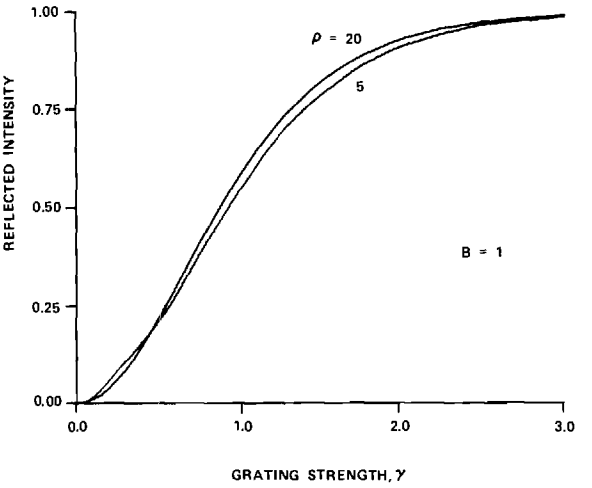


Fig. 2. Calculated intensity of the diffracted (reflected) wave for incidence at the first Bragg condition ($B = 1$).

In the next section the intensities of the transmitted and diffracted waves are calculated using the present exact coupled-wave analysis; these intensities are then compared to the approximate results of Kogelnik,[13] of Magnusson and Gaylord,[14] and of Bergstein and Kermisch.[5]

IV. Calculated Results and Discussion

Following the procedure outlined in Sec. II, the intensities of the reflected and transmitted waves are calculated for the first Bragg incidence $B = 1$, over a range of grating strengths γ from zero to 3.0, and for several values of the parameter ρ. The average permittivities of all three regions are taken to be equal so that comparisons can be made with the previous approximate coupled-wave analyses.[13,14] In Fig. 2 the calculated intensities of the reflected wave are plotted vs the grating strength γ for two different values of the parameter ρ. The transmitted and reflected intensities sum to unity. The intensity of the reflected wave increases initially in a parabolic manner and finally reaches saturation at a maximum value of 1 as γ approaches infinity. This corresponds to 100% conversion of the incident wave into the reflected wave. Nearly 100% conversion is possible with moderate grating strengths, for example, more than 97% conversion occurs at $\gamma = 2.5$. The errors in the calculated diffracted intensities due to the approximations in the Kogelnik[13] (two-wave), in the Magnusson and Gaylord[14] (multiwave) coupled-wave analyses, and in the Bergstein and Kermisch[5] modal analysis are plotted in Fig. 3. The absolute error is defined as the difference between the actual reflected wave intensity and the intensity calculated using each of the three approximate methods.

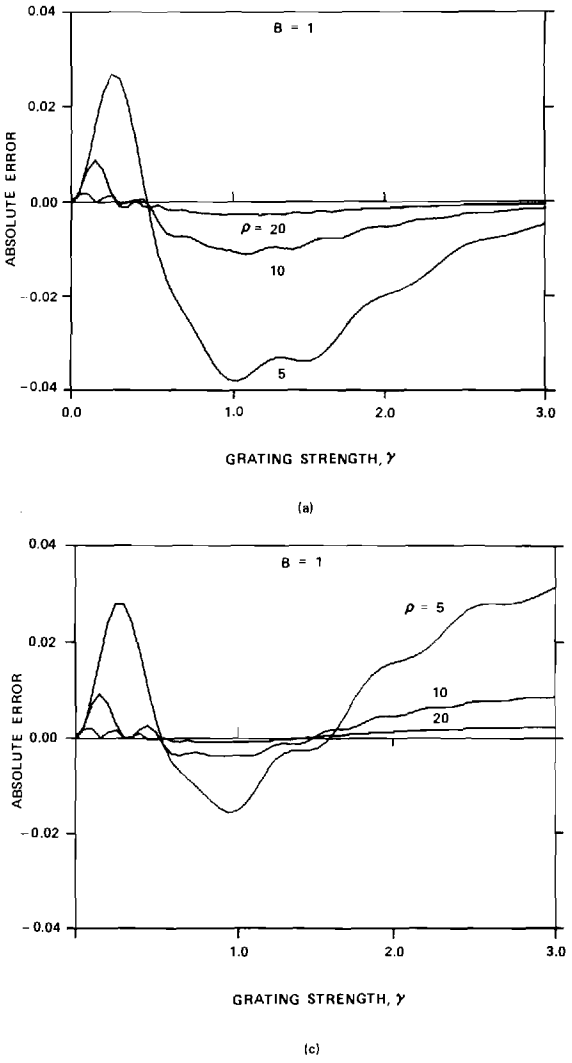


Fig. 3. The absolute error in the diffraction efficiency of the diffracted (reflected) wave defined as the difference between the diffraction efficiency calculated using the present exact analysis and the diffraction efficiency calculated using various approximate theories: (a) Kogelnik's two-wave coupled-wave analysis; (b) Magnusson and Gaylord's multiwave coupled-wave analysis; and (c) Bergstein and Kermisch's two-wave modal analysis.

Figure 3 shows that both the multiwave coupled-wave analysis of Magnusson and Gaylord,[14] and the two-wave modal analysis of Bergstein and Kermisch[5] produce roughly the same error and the Kogelnik[13] two-wave analysis produces slightly more error. This indicates that the second derivatives of the field amplitudes are as important as the higher-order waves for reflection gratings. Close examination of Fig. 3 shows that the error is usually within $\pm\rho^{-2}$. For example, for $\rho = 20$ the maximum magnitude of the error is 0.0027 [Fig. 3(a)], 0.0020 [Fig. 3(b)], and 0.0024 [Fig. 3(c)] respectively, corresponding to $\rho^{-2} = 0.0025$. For $\rho = 10$, the error is 0.011 [Fig. 3(a)], 0.009 [Fig. 3(b)], and 0.010 [Fig. 3(c)] respectively, corresponding to $\rho^{-2} = 0.010$. A large number of additional calculations confirm this general observation. Thus, diffraction efficiencies calculated from the approximate coupled-wave and modal analyses are essentially within $\pm\rho^{-2}$ of the exact results.

The reflected intensity for incidence at the second Bragg condition is plotted vs γ in Fig. 4 for several values of ρ. Figure 4 shows that for moderate values of the grating strength the reflected intensity is very small for

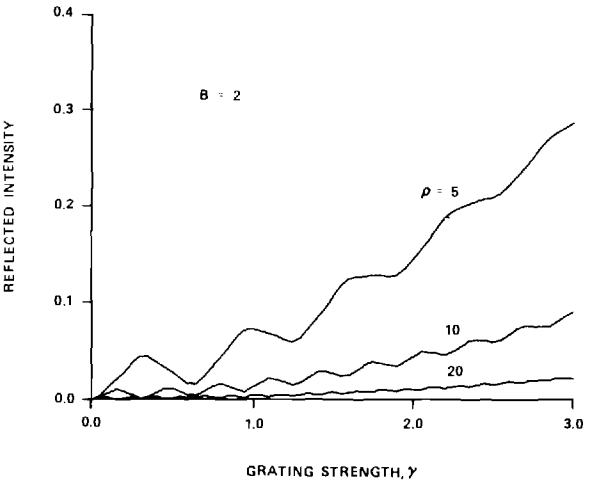


Fig. 4. Calculated diffraction efficiency of the diffracted (reflected) wave for incidence at the second Bragg condition ($B = 2$).

large ρ and that smaller values of ρ are needed to produce significant second-order diffraction. This indicates that ρ, which has been shown to delineate the two-wave diffraction regime in transmission gratings, plays the same role for reflection gratings.

V. Conclusion

The diffraction of a light wave by planar reflection gratings bounded by two different media has been analyzed using the recently-developed exact rigorous coupled-wave approach. The solution is cast in a simple matrix form that is suitable for straightforward solution using a digital computer. The present analysis is compared with previous approximate coupled-wave and modal analyses of reflection gratings. It is shown that the approximations introduced in previous work cause significant errors in the calculated results and cause the formulation to break down for light incident at even-order Bragg conditions. It is also shown that the error in the diffraction efficiency introduced by the various approximations in the previous coupled-wave and modal analyses is within $\pm 1/\rho^2$ of the exact value, where ρ is the regime parameter. Calculated results for incidence at higher-order Bragg conditions show that the reflected intensity is very small unless the regime parameter ρ is small. The present analysis may be applied to grating filters, distributed Bragg reflector lasers, and other devices that use reflection gratings.

This work was sponsored by the National Science Foundation under Grant ECS-7919592 and by the Joint Services Electronics Program under Grant DAAG29-78-C-0005.

References

1. T. Tamir, H. C. Wang, and A. A. Oliner, IEEE Trans. Microwave Theory Tech. **MTT-12,** 323 (1964).
2. T. Tamir and H. C. Wang, Can. J. Phys. **44,** 2073 (1966).
3. T. Tamir, Can. J. Phys. **44,** 2461 (1966).
4. C. B. Burckhardt, J. Opt. Soc. Am. **56,** 1502 (1966).
5. L. Bergstein and D. Kermisch, Proc. Symp. Modern Opt. **17,** 655 (1967).
6. R. S. Chu and T. Tamir, IEEE Trans. Microwave Theory Tech. **MTT-18,** 486 (1970).
7. R. S. Chu and T. Tamir, Proc. IEE **119,** 797 (1972).
8. F. G. Kaspar, J. Opt. Soc. Am. **63,** 37 (1973).
9. R.S. Chu and J. A. Kong, IEEE Trans. Microwave Theory Tech. **MTT-25,** 18 (1977).
10. S. T. Peng and E. S. Cassedy, Proc. Symp. Modern Opt. **17,** 299 (1967).
11. P. Phariseau, Proc. Indian Acad. Sci. Sect. A. **44,** 165 (1965).
12. W. R. Klein and B. D. Cook, IEEE Trans. Sonic Ultrason. **SU-14,** 123 (1967).
13. H. Kogelnik, Bell Sys. Tech. J. **48,** 2909 (1969).
14. R. Magnusson and T. K. Gaylord, J. Opt. Soc. Am. **67,** 1165 (1977).
15. J. A. Kong, J. Opt. Soc. Am. **67,** 825 (1977).
16. M. G. Moharam and T. K. Gaylord, J. Opt. Soc. Am., (1981) submitted.
17. M. G. Moharam, T. K. Gaylord, and R. Magnusson, Opt. Commun. **32,** 14 (1980).

Chain-matrix analysis of arbitrary-thickness dielectric reflection gratings

M. G. Moharam and T. K. Gaylord

School of Electrical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

Received August 12, 1981

A simple but rigorous chain-matrix analysis is used to determine the plane-wave diffraction efficiencies and angular selectivities of planar pure-reflection gratings (zero-slant angle) of arbitrary thickness, arbitrary-starting and arbitrary-ending conditions, and arbitrary incidence angle. The results from rigorous coupled-wave theory in the angular limit as the fringes become parallel to the surfaces are shown to approach these results in a particular average sense.

INTRODUCTION

Pure slab-reflection gratings have fringes parallel to the surfaces of the medium. It is due to their finite thicknesses that these structures are not strictly periodic. There may be a large number of grating periods present, but modal[1-8] and coupled-wave[9-15] grating theories based on the Floquet theorem require a truly periodic grating (an infinite number of periods). These theories may be applied in the angular limit as the slanted fringes of a general grating approach being parallel to the surface. However, for exact parallelism with the surface, a continuum of solutions is possible, depending on the number of periods, the starting conditions, and the ending conditions of the grating.

The chain-matrix method of solution[16,17] is often applied to the analysis of single-layer and multilayer thin-film coatings. However, a reflection grating with a continuous variation in permittivity may be represented as a large stack of thin homogeneous slabs.[18] If the individual planar slabs are allowed to be sufficiently thin, the resulting model can be made arbitrarily accurate. The chain matrix relates quantities at the input to the corresponding quantities at the output. In various forms, the chain matrix is also referred to as the characteristic matrix, the *ABCD* matrix, and the transmission matrix. This method of analysis is used here. It is rigorous for the case of pure-reflection gratings.

The purpose of this paper is (1) to present rigorously calculated plane-wave diffraction efficiency and angular selectivity results for dielectric reflection gratings showing the effect of grating thickness, the effect of fractional periods, and the effect of starting and ending conditions, and (2) to show that results from rigorous periodic-based theories approach, in the reflection grating limit, the average of the continuum of results from rigorous reflection grating theory.

THEORY

An obliquely incident plane wave illuminates a lossless dielectric reflection grating. The modulated region (region 2) extends from $z = 0$ to $z = d$. The regions outside the grating (regions 1 and 3) are homogeneous. The relative permittivity (dielectric constant) in the grating is of a general sinusoidal form given by

$$\epsilon(z) = \epsilon_2 + \Delta\epsilon \cos(Kz - \Phi), \qquad 0 \le z \le d, \tag{1}$$

where ϵ_2 is the average relative permittivity of region 2, $\Delta\epsilon$ is the amplitude of the grating, $K = 2\pi/\Lambda$, Λ is the grating period, and Φ is a phase constant that determines the starting condition for the grating ($\Phi = 0, \pi/2, \pi, 3\pi/2$ correspond, respectively, to cosine, sine, -cosine, -sine gratings). For this case of a planar-reflection grating with fringes exactly parallel to the surfaces, there are no higher-order waves: that is, the incident wave (region 1) is divided into a transmitted wave (region 3) and a diffracted (reflected) wave (region 1). There are no other waves for this grating configuration.

The planar-reflection grating given by Eq. (1) may be represented by a large number of sufficiently thin homogeneous slabs. The incident and reflected electric fields at the input of the ith slab may be related to those at the input of the $(i+1)$th slab by a chain matrix

$$\begin{pmatrix} E_{\text{inc},i} \\ E_{\text{refl},i} \end{pmatrix} = \begin{pmatrix} A_i & B_i \\ B_i^* & A_i^* \end{pmatrix} \begin{pmatrix} E_{\text{inc},i+1} \\ E_{\text{refl},i+1} \end{pmatrix}. \tag{2}$$

In this type of formulation, the effect of having multiple slab layers together is represented simply by multiplying their respective chain matrices together and then multiplying that product by the chain matrix of region 3 ($z > d$). Thus, for a reflection grating that is divided into n thin slabs, the governing equation is

$$\begin{pmatrix} 1 \\ E_r \end{pmatrix} = \prod_{i=1}^{n} \begin{pmatrix} A_i & B_i \\ B_i^* & A_i^* \end{pmatrix} \times \frac{1}{\tau_{n+1}} \begin{pmatrix} 1 & r_{n+1} \\ r_{n+1} & 1 \end{pmatrix} \begin{pmatrix} E_t \\ 0 \end{pmatrix}, \tag{3}$$

where E_r and E_t are the reflected and transmitted normalized electric-field amplitudes, respectively. For a lossless dielectric layer, the matrix elements are

$$A_i = (1/\tau_i)\exp[j2\pi(\epsilon_{2,i})^{1/2} d_i \cos\theta_{2,i}/\lambda], \tag{4}$$

$$B_i = r_i A_i^*, \tag{5}$$

where λ is the free-space wavelength and $\epsilon_{2,i}$, $\theta_{2,i}$, and d_i are the relative permittivity, the angle of refraction, and the thickness of the ith layer. The amplitude reflection and

transmission coefficients are

$$r_i = (Z_i - Z_{i-1})/(Z_i + Z_{i-1}),$$
$$\tau_i = 1 + r_i, \tag{6}$$

where Z_i is the normalized characteristic impedance of the ith layer. For H-mode polarization (electric field perpendicular to plane of incidence),

$$Z_0 = 1, \quad \text{region 1},$$
$$Z_i = \cos\theta/[(\epsilon_{2,i}/\epsilon_1) - \sin^2\theta]^{1/2}, \quad 1 \le i \le n,$$
$$Z_{n+1} = \cos\theta/[(\epsilon_3/\epsilon_1) - \sin^2\theta]^{1/2}, \quad \text{region 3}, \tag{7}$$

and for E-mode polarization (electric field parallel to plane of incidence),

$$Z_0 = 1, \quad \text{region 1},$$
$$Z_i = (\epsilon_1/\epsilon_{2,i})[(\epsilon_{2,i}/\epsilon_1) - \sin^2\theta]^{1/2}/\cos\theta, \quad 1 \le i \le n,$$
$$Z_{n+1} = (\epsilon_1/\epsilon_3)[(\epsilon_3/\epsilon_1) - \sin^2\theta]^{1/2}/\cos\theta, \quad \text{region 3}, \tag{8}$$

where θ is the angle of incidence in region 1. The diffraction efficiency (fraction of power reflected) is simply $DE = E_r E_r^*$.

RESULTS AND DISCUSSION

Since much of the literature describes the simple case of the same average permittivity inside and outside the grating ($\epsilon_1 = \epsilon_2 = \epsilon_3$), the situation is also used here. This case applies directly, for example, in integrated optics. However, this is not an essential assumption, and the theory applies in general.

At the input surface ($z = 0$), the grating may start with any value of relative permittivity. The permittivity may be continuous at $z = 0$, as in the cases of the +sine and −sine gratings, or may be discontinuous by a maximum amount, as in the +cosine and −cosine gratings. The starting condition for the grating is determined by the value of Φ in Eq. (1). There is a continuum of starting conditions given by the range $0 \le \Phi \le 2\pi$.

Typical reflected intensities are shown in Fig. 1 for H-mode polarization for a thickness of 3.125 grating periods and $\rho = 20$, where ρ, the Bragg-regime parameter, is given by $\rho = 2\lambda^2/\Lambda^2\Delta\epsilon$. The input wave is incident at the Bragg angle. That is, the Bragg condition $\lambda/(\epsilon_1)^{1/2} = 2\Lambda\cos\theta$ is satisfied. Thus, for the present case, $\rho = 8\cos^2\theta(\epsilon_1/\Delta\epsilon)$. For the particular grating thickness and modulation of Fig. 1, the $\Phi = 9\pi/8$ grating (between a −cosine grating and a −sine grating) has the largest diffraction efficiency (58.8%). For the same thickness and modulation the $\Phi = 5\pi/8$ grating (between a +sine grating and a −cosine grating) and the $\Phi = 13\pi/8$ grating (between a −sine grating and a +cosine grating) both have the lowest diffraction efficiency (54.3%). Gratings with all other starting conditions (values of Φ) have efficiencies between these two values. This continuum of solutions is represented by the vertical line in Fig. 1. For any thickness and modulation, the maximum diffraction efficiency is found to occur for the grating with

$$\Phi = [1 + (d/\Lambda)\text{modulo } 1]\pi \tag{9}$$

for $0 < (d/\Lambda)\text{modulo } 1 \le 0.5$ and

$$\Phi = [\pm(1/2) + (d/\Lambda)\text{modulo } 1]\pi \tag{10}$$

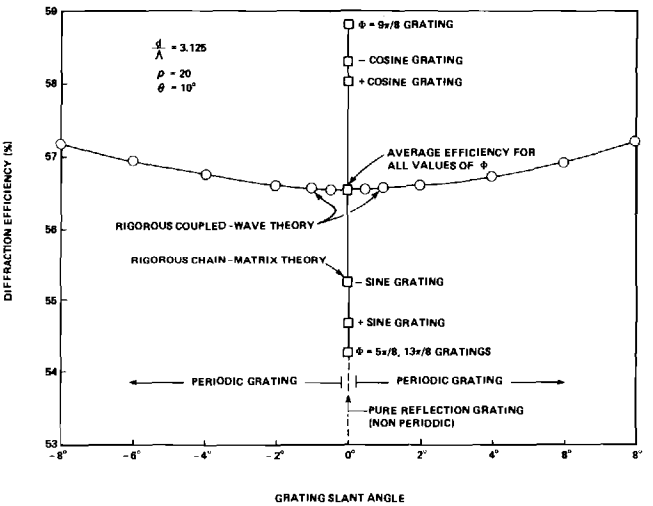


Fig. 1. Rigorously calculated total on-Bragg-angle diffraction efficiencies for H-mode polarization for dielectric reflection gratings with $\epsilon_1 = \epsilon_2 = \epsilon_3$. The results from rigorous coupled-wave theory for slanted reflection gratings are shown to approach the average of the continuum of solutions for the unslanted pure-reflection gratings.

for $0.5 < (d/\Lambda)\text{modulo } 1 \le 1.0$. For example, if $d/\Lambda = 5.5$, the maximum efficiency occurs for $\Phi = 3\pi/2$ (a −cosine grating). If $d/\Lambda = 7.75$, the maximum efficiency occurs for gratings with $\Phi = \pi/4$ and $\Phi = 5\pi/4$. The ending condition for a grating is determined by its thickness. By the reciprocity principle the +sine and −sine gratings produce the same efficiency for thicknesses that are an integer number of periods. Likewise, the +cosine and −cosine gratings have the same efficiency for thicknesses that are an odd-integer multiple of a half-period.

Also plotted in Fig. 1 is the total diffracted intensity as calculated by rigorous coupled-wave theory. This theory applies to periodic gratings with slanted fringes. The grating slant angle may approach (but not equal) zero in this theory. A grating slant angle of zero corresponds to a nonperiodic pure-reflection grating. The transition of the grating slant angle from near zero to exactly zero involves a singularity. The total diffraction efficiency changes from a single-valued function to a multiple-valued function (a continuum of solutions), as shown in Fig. 1. An important result of this work is that the rigorous coupled-wave theory in the limit of vanishingly small slant angle predicts a total diffraction efficiency equal to the average of the pure-reflection grating efficiencies with respect to the entire range of starting conditions ($0 \le \Phi \le 2\pi$). That is, for a given thickness and given modulation pure-reflection grating, the average diffraction efficiency for all possible starting conditions is equal to the diffraction efficiency for a slightly slanted reflection grating. In the case of a slanted-fringe reflection grating, the permittivity is periodic along the grating surfaces ($z = 0, d$) even if the slant is small. Only for the exactly unslanted case is the permittivity equal to a constant over all the grating surfaces. It seems intuitively correct that a slightly slanted reflection grating behaves like the average of all possible reflection gratings of the same thickness and modulation. The slanted grating is periodic along the $z = 0$ surface, at some points starting at a peak (like a cosine grating), at some points starting at a minimum (like a −cosine grating), etc.

The diffraction efficiencies of the +cosine and +sine gratings as calculated by the rigorous chain-matrix theory for H-mode polarization for the cases of $\epsilon_1 = \epsilon_2 = \epsilon_3$ are shown in Fig. 2 as a deviation from the average value for all possible starting conditions. This average value is within $\pm 1/\rho^2$ (which in this case is ±0.25%) of the diffraction efficiency predicted by $DE = \tanh^2[\pi \Delta \epsilon d/2\lambda(\epsilon_1)^{1/2} \cos\theta]$ from Kogelnik.[9] As can be seen from this figure, a grating thickness that is a multiple of $\Lambda/4$ produces a small (or zero) deviation of the diffraction efficiency from the average. The largest deviations from the average occur for thicknesses that are odd multiples of $\Lambda/8$. A plot of the deviations for the −cosine and −sine gratings is similar to Fig. 2 and has all the basic features of Fig. 2. When the grating has a thickness that is an integer multiple of Λ, the diffraction efficiencies of the ±sine gratings are exactly the same. For this case, there are small deviations in efficiency between the +cosine grating, the −cosine grating, and the ±sine gratings. As the thickness of the reflection grating increases, the deviations from the average diffraction efficiency, as shown in Fig. 2, increase as d/Λ increases from zero to about 3, and then the deviation oscillations decrease in amplitude as d/Λ increases further. For this case of $\rho = 20$ and for a grating thickness greater than 10 periods, the presence of fractional periods produces essentially negligible variations from the case of integer period thickness. The diffraction efficiency deviation from the average value (or equivalently the deviation from the rigorous coupled-wave theory result) does not exceed approximately ±2.5% efficiency for any thickness. Thus the results from rigorous coupled-wave theory with the grating slant angle approaching zero are within ±2.5% of the rigorous pure-reflection grating results regardless of the thickness and the starting conditions. In addition, the average value of the diffraction efficiency for all starting conditions is within $\pm 1/\rho^2$ of the value predicted by Kogelnik's theory.[9] Therefore, in this case, the deviations (±2.5%) that are due to starting conditions are an order of magnitude greater than the error (±0.25%) that is due to the neglect of second derivatives of field amplitudes and neglect of boundary diffraction in the Kogelnik coupled-wave formulation.

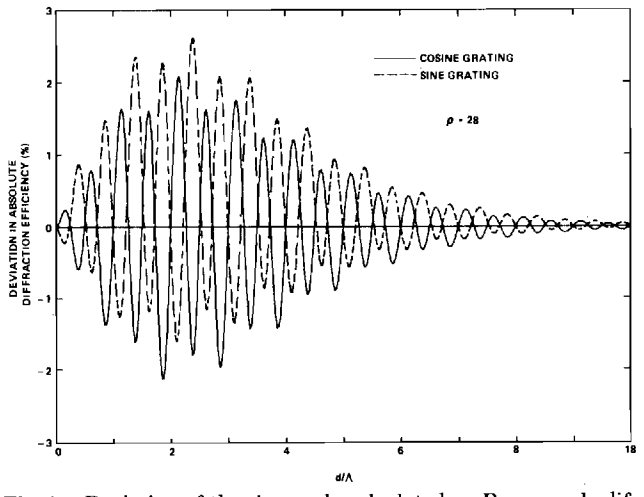


Fig. 2. Deviation of the rigorously calculated on-Bragg-angle diffraction efficiencies for H-mode polarization for +cosine and +sine gratings from the average of the continuum of diffraction efficiencies for pure-reflection gratings as a function of thickness.

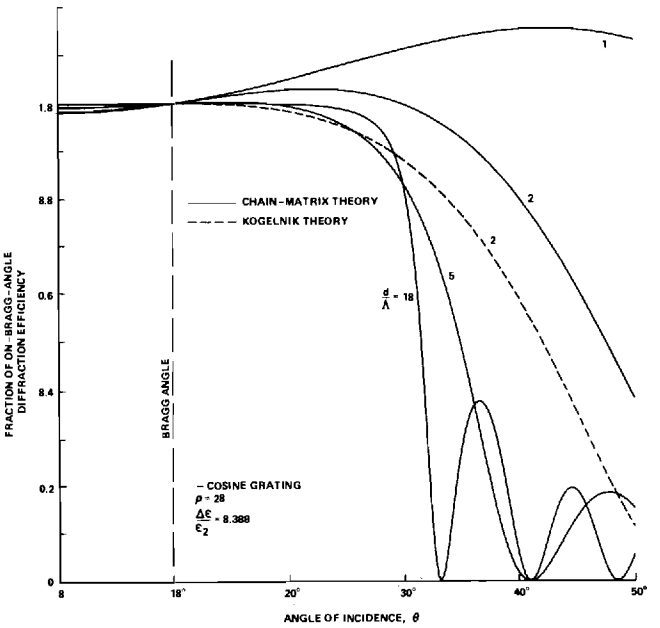


Fig. 3. Angular selectivity characteristics (fraction of on-Bragg-angle diffracted power that is reflected as a function of angle of incidence) for H-mode polarization for pure-reflection gratings of various thicknesses.

Typical plots of chain-matrix-calculated angular selectivity responses of pure-reflection gratings (−cosine gratings) are shown in Fig. 3. The diffraction efficiency is normalized to the value at the Bragg angle (10° in this example case). The angular selectivity is symmetric about normal incidence. The diffracted (reflected) power remains nearly constant from normal incidence to the Bragg angle. Beyond the Bragg angle the diffracted power decreases, thus producing the well-known angular selectivity feature of thick gratings.[9] For gratings of only two periods' thickness, the large Fresnel-reflection component may cause the total reflected power to increase initially for angles greater than the Bragg angle. As the thickness of the grating increases, the angular selectivity is enhanced. That is, to achieve a large fraction of reflected power, the input wave must be incident in an angular corridor of decreasing width as the grating thickness increases. For comparison, the corresponding angular selectivity from Kogelnik's approximate two-wave coupled-wave theory is also plotted for the case of a grating that is two periods thick. This approximate theory predicts a smaller angular selectivity width than that from rigorous theory. The approximate Kogelnik theory also does not predict the initial rise in reflected power beyond the Bragg angle that is due to the Fresnel component of the reflection. Angular selectivity plots for ±sine and +cosine gratings are similar in general form to the −cosine grating case shown in Fig. 3.

SUMMARY

A rigorous chain-matrix method of analysis has been used to calculate the diffraction characteristics of pure- (unslanted) reflection gratings. Since pure-reflection gratings diffract (reflect) only a single order, the diffraction efficiency of this order is a complete description of the diffraction process. The analysis has been applied to reflection gratings of arbitrary

thickness, arbitrary starting and ending conditions, and arbitrary incidence angle. For the same average permittivity inside and outside the grating, a variation in the starting conditions of the grating (e.g., ±cosine and ±sine grating) produces approximately ±2.5% maximum change in the diffraction efficiency from the average value for a grating with $\rho = 20$. The results from rigorous coupled-wave theory, in the angular limit of the fringes' becoming parallel to the surface, have been shown to approach the average value of the pure-reflection grating diffraction efficiencies for all possible starting conditions.

The rigorous chain-matrix method has been applied here to sinusoidal grating profiles with the same average permittivity inside and outside the grating ($\epsilon_1 = \epsilon_2 = \epsilon_3$). However, other grating profiles and other surrounding permittivities may be treated by using the analysis presented without modification.

REFERENCES

1. C. B. Burckhardt, "Diffraction of a plane wave at a sinusoidally stratified dielectric grating," J. Opt. Soc. Am. **56,** 1502–1509 (1966).
2. T. Tamir and H. C. Wang, "Scattering of electromagnetic waves by a sinusoidally stratified half-space: I. Formal solution and analytic approximations," Can. J. Phys. **44,** 2073–2094 (1966).
3. T. Tamir, "Scattering of electromagnetic waves by a sinusoidally stratified half-space: II. Diffraction aspects at the Rayleigh and Bragg wavelengths," Can. J. Phys. **44,** 2461–2494 (1966).
4. L. Bergstein and D. Kermisch, "Image storage and reconstruction in volume holography," Proc. Symp. Modern Opt. **17,** 655–680 (1967).
5. S. T. Peng and E. S. Cassedy, "Scattering of light waves at boundaries to parametrically modulated media," Proc. Symp. Modern Opt. **17,** 299–342 (1967).
6. F. G. Kaspar, "Diffraction by thick periodically stratified gratings with complex dielectric constant," J. Opt. Soc. Am. **63,** 37–45 (1973).
7. R. S. Chu and T. Tamir, "Wave propagation and dispersion in space-time periodic media," Proc. IEE **119,** 797–806 (1972).
8. R. S. Chu and J. A. Kong, "Modal theory of spatially periodic media," IEEE Trans. Microwave Theory Tech. **MTT-25,** 18–24 (1977).
9. H. Kogelnik, "Coupled wave theory for thick hologram gratings," Bell Syst. Tech. J. **48,** 2909–2947 (1969).
10. R. S. Chu and T. Tamir, "Guided-wave theory of light diffraction by acoustic microwaves," IEEE Trans. Microwave Theory Tech. **MTT-18,** 486–504 (1970).
11. D. L. Jaggard and C. Elachi, "Floquet and coupled-wave analysis of higher-order Bragg coupling in a periodic medium," J. Opt. Soc. Am. **66,** 674–682 (1976).
12. W. R. Klein and B. D. Cook, "Unified approach to ultrasonic light diffraction," IEEE Trans. Sonics Ultrason. **SU-14,** 123–134 (1967).
13. R. Magnusson and T. K. Gaylord, "Analysis of multiwave diffraction of thick gratings," J. Opt. Soc. Am. **67,** 1165–1170 (1977).
14. M. G. Moharam and T. K. Gaylord, "Coupled-wave analysis of reflection gratings," Appl. Opt. **20,** 240–244 (1981).
15. M. G. Moharam and T. K. Gaylord, "Rigorous coupled-wave analysis of planar grating diffraction," J. Opt. Soc. Am. **71,** 811–818 (1981).
16. M. Born and E. Wolf, *Principles of Optics*, 6th ed. (Pergamon, Oxford, 1980), pp. 55–70.
17. R. E. Collin, *Field Theory of Guided Waves* (McGraw-Hill, New York, 1960), pp. 79–94.
18. W. W. Rigrod, "Diffraction efficiency of nonsinusoidal Bragg reflection gratings," J. Opt. Soc. Am. **64,** 97–99 (1974); Erratum, J. Opt. Soc. Am. **64,** 895 (1974).

To be published
J. Opt. Soc. Am.
October 1982

DIFFRACTION ANALYSIS OF DIELECTRIC SURFACE-RELIEF GRATINGS

M. G. Moharam and T. K. Gaylord
School of Electrical Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332

Abstract

Diffraction by a dielectric surface-relief grating is analyzed using rigorous coupled-wave theory. The analysis applies to arbitrary grating profiles, groove depths, angles of incidence, and wavelengths. Example results for a wide range of groove depths are presented for sinusoidal, square-wave, triangular, and sawtooth gratings. Diffraction efficiencies obtained from the present method of analysis are compared with previously published numerical results. To obtain large diffraction efficiencies (greater than 85%) for gratings with typical substrate permittivities, it is shown that the grating profile should be expressable as an even function.

1. INTRODUCTION

Dielectric surface-relief (or corrugated) gratings are of wide interest owing to their many applications in quantum electronics, integrated optics, spectroscopy, and holography.[1] Example devices include distributed feedback lasers, distributed Bragg reflector lasers, beam deflectors, waveguide couplers, spectral filters, wavelength multiplexers and demultiplexers, beam deflectors, holographic beam combiners, and others. Several investigators have obtained rigorous solutions for the exact electromagnetic boundary value problem that apply to gratings with rectangular or triangular grooves.[2-8] Others have obtained approximate results by applying perturbation techniques.[9-10] For gratings with arbitrary profiles, the integral method[1,11-13] was first used to obtain numerical results. Later the differential method was developed[1,2] and it is probably the most widely-used method for analyzing arbitrary-profile gratings.[2,14-16] The differential method requires advanced numerical techniques such as the combined Runge-Kutta and Adam-Moulton algorithm,[1,2] the unimoment method,[14] Numerov's algorithm,[1] the modified Adam-Moulton algorithm,[16] or other similar algorithms. Some of these algorithms do not converge for relatively deep grooves.

Recently, a rigorous coupled-wave approach has been formulated and applied to planar dielectric gratings.[17] A state variables representation of the coupled-wave amplitudes allows closed-form expressions to be obtained for these amplitudes in terms of the eigenfunctions and eigenvectors of the coefficient matrix defined by the rigorous coupled-wave equations.

In this paper, the rigorous coupled-wave approach is adapted to the exact electromagnetic boundary value problem associated with dielectric surface-relief gratings of arbitrary profile. In this method, a surface-relief grating is divided into a large number of thin layers parallel to the surface. Each thin grating is then analyzed using the state variables method of solution of the rigorous coupled-wave equations for that grating. By formulating the problem in a particular manner, it is shown that the grating layers may be treated one-at-a-time in sequence thus reducing the numerical calculations to an easily manageable size. There are no approximations in the analysis and results are obtainable to an arbitrary level of accuracy. The diffraction efficiencies of all orders of both the transmitted and reflected waves are determined in the process. Peng, Tamir, and Bertoni[5] were the first to propose partitioning the grating into thin layers and approximating each layer by a rectangular profile. They then calculated the field in each layer using the Lewis and Hessel[18] method of analyzing a periodic array of slabs. However, they presented numerical results only for one[3] and two-layered[4] gratings. In the present work, the rigorous coupled-wave approach is used to find the fields in each layer and a large number of layers (typically 50 to 100) are used.

To illustrate the utility of this method of analysis, the diffraction characteristics of sinusoidal, square-wave, triangular, and sawtooth gratings are determined for a wide range of groove depths for a plane wave incident at the first Bragg angle. Very large groove depths (four grating periods deep for the examples presented) may be handled routinely with this analysis without any type of numerical difficulties. In a final section the accuracy and the convergence of the present method is discussed in detail. The diffraction efficiency results obtained by the present

method are compared to numerical results obtained by other techniques.[12-14,16] It is shown that the rigorous coupled-wave method accurately predicts conservation of power when all of the individual diffracted powers are summed.

2. THEORY

The general dielectric surface-relief grating diffraction problem treated in this paper is shown in Fig. 1. An electromagnetic plane wave is obliquely incident upon the grating. The periodic structure produces, in general, both forward-diffracted and backward-diffracted waves as shown in the figure. Region 1 (the input region) is a homogeneous dielectric with a relative permittivity (dielectric constant) of ε_I. Likewise, region 3 is homogeneous with a relative permittivity of ε_{III}. Region 2 (the grating region) consists of a periodic distribution of both types of dielectrics. The boundary between the ε_I dielectric and the ε_{III} dielectric in region 2 is given by

$$z = F(x) = F(x + \Lambda) \tag{1}$$

where Λ is the grating period. The function $F(x)$ thus represents the grating surface profile. There are no restrictions on the form of $F(x)$ in this analysis. Curved lines, straight lines, shadow regions, etc. are all allowed. In this paper, for simplicity, the case of lossless dielectrics and incident plane wave polarization perpendicular to the plane of incidence (H-mode) are treated.

The total electric field in region 1 is the sum of the incident and the backward-traveling waves. The normalized total electric field in region 1 may be expressed as

$$E_1 = \exp(-j\bar{k}_1 \cdot \bar{r}) + \sum_{i=-\infty}^{\infty} R_i \exp(-j\bar{k}_{1i} \cdot \bar{r}) \tag{2}$$

where $j = (-1)^{1/2}$, $\bar{k}_1$ is the incident field wavevector of magnitude, $k_1 = 2\pi(\varepsilon_I)^{1/2}/\lambda$, λ is the free space wavelength, and $\bar{r} = x\hat{x} + y\hat{y} + z\hat{z}$ with $\hat{}$ denoting unit vectors. R_i is the normalized amplitude of the i-th reflected wave in region 1 with wavevector $\bar{k}_{1i}$. Likewise the normalized total electric field in region 3 is

$$E_3 = \sum_{i=-\infty}^{\infty} T_i \exp[-j\bar{k}_{3i} \cdot (\bar{r} - d\hat{z})] \tag{3}$$

where T_i is the normalized amplitude of the i-th transmitted wave into region 3 with wavevector $\bar{k}_{3i}$ and d is the groove depth. The quantities $\bar{k}_{1i}$ and $\bar{k}_{3i}$ are determined below by using the phase matching requirement.

In the present analysis, the grating region (region 2) is divided into N thin planar grating slabs perpendicular to the z-axis. Then the rigorous coupled-wave analysis that has been developed for planar grating is applied to each slab grating.[17] If the individual planar gratings are sufficiently thin, any grating profile can be analyzed to an arbitrary level of accuracy. The n-th slab within region 2 as shown in Fig. 2 will consist of a periodic distribution of ε_I and ε_{III} dielectrics. The relative permittivity for the n-th slab grating is periodic, $\varepsilon_n(x,z_n) = \varepsilon_n(x + \Lambda,z_n)$, and may be expanded in a Fourier series as

$$\varepsilon_n(x, z_n) = \varepsilon_I + (\varepsilon_{III} - \varepsilon_I) \sum_{h=-\infty}^{+\infty} \tilde{\varepsilon}_{h,n} \exp(jhKx) \tag{4}$$

where z_n is the z-coordinate of the n-th slab, h is the harmonic index, K is the magnitude of the grating vector ($K = 2\pi/\Lambda$), and $\tilde{\varepsilon}_{h,n}$ are the normalized complex harmonic amplitude coefficients given by

$$\tilde{\varepsilon}_{h,n} = (1/\Lambda) \int_0^{\Lambda} f(x, z_n) \exp(-jhKx)\, dx \tag{5}$$

where the function $f(x,z_n)$ is equal to either zero or unity depending whether, for a particular value of x, the grating relative permittivity is ε_I or ε_{III} respectively.

In the coupled-wave representation used in this treatment, the fields inside each grating slab are expanded in terms of the space harmonics of the fields in the periodic structure. These space harmonics inside the grating correspond to diffracted orders outside of the grating. Thus, the partial fields inside the modulated medium are visualized as diffracted waves that progress through the planar slab and couple energy back and forth between each other as they progress. In the coupled-wave approach the total field is thus expressed as

$$E_{2,n} = \sum_{i=-\infty}^{+\infty} S_{i,n}(z) \exp(-j\bar{\sigma}_{i,n} \cdot \bar{r}) \tag{6}$$

where i is the space harmonic index (integer) and $S_{i,n}(z)$ are the space harmonic field amplitudes. As a result of the Floquet theorem, the diffracted wavevectors inside the n-th grating, $\bar{\sigma}_{i,n}$, may be represented

$$\bar{\sigma}_{i,n} = \bar{k}_{2,n} - i\bar{K} \tag{7}$$

where $\bar{K} = K\hat{x}$ and $\bar{k}_{2,n}$ is the wavevector of the zero-order (i=0) refracted wave having a magnitude of $k_{2,n} = 2\pi(\varepsilon_{0,n})^{\frac{1}{2}}/\lambda$, and $\varepsilon_{0,n}$ is the average relative permittivity for the n-th slab grating.

Each i-th diffracted field in regions 1 and 3 must be phase matched to the i-th space harmonic field inside each n-th slab grating. Thus $\bar{k}_{1i} \cdot \hat{x} = (\bar{k}_{2,n} - i\bar{K}) \cdot \hat{x} = \bar{k}_{3i} \cdot \hat{x}$ for any i and n. In the homogeneous regions (1 and 3) the backward- and forward-diffracted waves have wavevectors with magnitudes $|\bar{k}_{1i}| = |\bar{k}_1|$ and $|\bar{k}_{3i}| = |\bar{k}_3|$ where $k_3 = 2\pi(\varepsilon_{III})^{\frac{1}{2}}/\lambda$. Knowing the total amplitudes and the x components of the diffracted wavevectors, the z components are then determined to be $\bar{k}_{1i} \cdot \hat{z} = [|\bar{k}_1|^2 - (\bar{k}_{1i} \cdot \hat{x})^2]^{\frac{1}{2}}$ and $\bar{k}_{3i} \cdot \hat{z} = [|\bar{k}_3|^2 - (\bar{k}_{3i} \cdot \hat{x})^2]^{\frac{1}{2}}$. Therefore the total fields in all regions may be rewritten as

$$E_1 = \exp\{-j[k_1(\sin\theta' \, x + \cos\theta' \, z)]\} + \sum_{i=-\infty}^{+\infty} R_i \exp\{-j[(k_1 \sin\theta' - iK) \, x + [k_1^2 - (k_1 \sin\theta' - iK)^2]^{\frac{1}{2}} z]\} \quad , \tag{8}$$

$$E_{2,n} = \sum_{i=-\infty}^{+\infty} S_{i,n}(z) \exp\{-j[(k_1 \sin\theta' - iK) \, x + (k_{2,n}^2 - k_1^2 \sin^2\theta')^{\frac{1}{2}} z]\} \quad , \tag{9}$$

and

$$E_3 = \sum_{i=-\infty}^{\infty} T_i \exp\{-j[(k_1 \sin\theta' - iK) \, x + [k_3^2 - (k_1 \sin\theta' - iK)^2]^{\frac{1}{2}} (z-d)]\} \quad . \tag{10}$$

In region 2, the wave equation for H-mode polarization

$$\nabla^2 E_{2,n} + k^2 \varepsilon_n(x, z_n)\, E_{2,n} = 0 \tag{11}$$

where $k = 2\pi/\lambda$, must be satisfied for each n-th slab grating. Substituting $\varepsilon_n(x,z_n)$ from Eq. (4) and $E_{2,n}$ from Eq. (9) into the wave equation, performing the indicated differentiations, and setting the coefficient of each exponential term equal to zero for nontrivial solutions yields the rigorous coupled-wave equations for the n-th slab grating:

$$\frac{d^2 S_{i,n}(z)}{dz^2} - j2(k_{2,n}^2 - k_1^2 \sin^2\theta')^{\frac{1}{2}} \frac{dS_{i,n}(z)}{dz} + K^2 i(m-i)\, S_{i,n}(z) + \\ + k^2 (\varepsilon_{III} - \varepsilon_I) \sum_{h=1}^{\infty} [\tilde{\varepsilon}_{h,n} S_{i-h,n}(z) + \tilde{\varepsilon}^*_{h,n} S_{i+h,n}(z)] = 0 \,. \tag{12}$$

These coupled-wave equations are an infinite set of second-order coupled difference-differential equations. Each diffracted wave (i) is coupled to other diffracted waves through the harmonics of the grating (i-h and i+h). The quantity m has been defined as

$$m = 2\Lambda\, (\varepsilon_I)^{\frac{1}{2}} \sin\theta'/\lambda \quad . \tag{13}$$

For the case when m is an integer, this represents a Bragg condition.

3. METHOD OF SOLUTION

The dielectric surface-relief grating diffraction problem as formulated in the previous section will be solved in a sequence of steps. First, the rigorous coupled-wave equations will be solved for the n-th slab

grating using a state variables method of solution. Second, electromagnetic boundary conditions (continuity of tangential E and tangential H) will be applied between region 1 and the first slab grating, then between the first and second slab gratings, and so forth and finally between the N-th slab grating and region 3. Third, the resulting array of boundary condition equations are solved for the reflected and transmitted diffracted amplitudes, R_i and T_i. From these amplitudes, the diffracted efficiencies are determined directly.

Using the methods of linear systems analysis[19] the coupled-wave differential equation description of the diffraction problem may be transformed into a state space description. Defining the state variables for the n-th slab grating as

$$S_{1,i,n}(z) = S_{i,n}(z) \tag{14}$$

$$S_{2,i,n}(z) = dS_{i,n}(z)/dz \tag{15}$$

transforms the infinite set of second-order differential equations (12) into two infinite sets of first-order state equations:

$$\frac{dS_{1,i,n}(z)}{dz} = S_{2,i,n}(z) \tag{16}$$

$$\begin{aligned}\frac{dS_{2,i,n}(z)}{dz} = &- k^2(\varepsilon_{III} - \varepsilon_I) \sum_{h=1}^{\infty} \tilde{\varepsilon}_{h,n}\, S_{1,i-h,n}(z) + \\ &- K^2 i(m-i)\, S_{1,i,n}(z) - k^2(\varepsilon_{III} - \varepsilon_I) \sum_{h=1}^{\infty} \tilde{\varepsilon}^*_{h,n} S_{1,i+h,n}(z) + \\ &+ j2(k_{2,n}^2 - k_1^2 \sin^2\theta')\, S_{2,i,n}(z) \quad .\end{aligned} \tag{17}$$

In matrix form, the state equation for the n-th slab grating may be written as:

$$\begin{bmatrix} \vdots \\ \dot{\tilde{S}}_{1,p,n} \\ \vdots \\ \dot{\tilde{S}}_{2,p,n} \\ \vdots \end{bmatrix} = \left[\begin{array}{c|c} a_{p,q,n} & b_{p,q,n} \\ \hline c_{p,q,n} & d_{p,q,n} \end{array}\right] \begin{bmatrix} \vdots \\ \tilde{S}_{1,q,n} \\ \vdots \\ \tilde{S}_{2,q,n} \\ \vdots \end{bmatrix} \tag{18}$$

where $\tilde{S}_{\ell,p,n} \equiv S_{\ell,i,n}$ (for ℓ=1,2), $\dot{S} = dS/dz$ and the elements of the four submatrices (p = 1 to s and q = 1 to s) are specified by Eqs. (16) and (17) for the n-th slab grating. The integers p and q are the row and column indices of the four submatrices. The maximum value of these indicies, s, is equal to the number of diffracted orders retained in the analysis. The value p = 1 corresponds to the most negative order (value of i) retained in the analysis and p = s corresponds to the most positive order retained. For example, if an odd number of waves are retained symmetrically about i=0 (the undiffracted wave) in the analysis, then p = i + (s+1)/2. Equation (18) corresponds to an unforced state equation and may be expressed concisely as $\dot{\mathbf{S}} = \mathbf{AS}$ where $\dot{\mathbf{S}}$ and $\mathbf{S}$ are the column vectors from Eq. (14) and $\mathbf{A}$ is the total coefficient matrix. Although $\mathbf{A}$ is an infinite matrix, results may be obtained in practice to an arbitrary level of accuracy with a truncated matrix. Each of the four submatrices is truncated to s x s. As the integer s increases, the calculated results rapidly converge to the exact results. The solutions of Eq. (18) are[19]

$$\tilde{S}_{p',n}(z) = \sum_{q'=1}^{2s} C_{q',n}\, w_{p',q',n} \exp(\lambda_{q',n} z) \tag{19}$$

where $\tilde{S}_{\ell,p,n}$ (for ℓ=1,2) has been rewritten as $\tilde{S}_{p',n}$ with $p' = p + (\ell-1)s$. The quantities $\lambda_{q',n}$ and $w_{p',q',n}$ are the eigenvalues and eigenvectors of the matrix **A**. These values are typically calculated using a computer library program.[20] The integers p' and q' are the row and column indices of the eigenvector matrix [w] and p' = 1 to 2s and q' = 1 to 2s. The quantities $C_{q',n}$ are unknown constants to be determined by the boundary conditions. The desired diffracted wave amplitudes for the n-th grating layer are given by

$$S_{i,n}(z) = \tilde{S}_{p',n}(z) \tag{20}$$

where p' is chosen to correspond to the i-th diffracted wave.

Electromagnetic boundary conditions require that the tangential electric and tangential magnetic fields be continuous across the boundaries between the slabs. For the H mode polarization described in this paper, the electric field only has a tangential component (y direction). The tangential component of H is in the x direction and from Maxwell's equations it is given by $H_x = (-j/\omega\mu)\, E_y/\partial z$. Therefore, for the boundary (z=0) between region 1 (the input region) and the first slab grating, the boundary condition for tangential E is

$$\delta_{io} + R_i = \sum_{q'=1}^{2s} C_{q',1}\, w_{p',q',1} \tag{21}$$

and for tangential H is

$$j(\bar{k}_{1i} \cdot \hat{z})(R_i - \delta_{i0}) = \sum_{q'=1}^{2s} C_{q',1}\, w_{p',q',1}\, [\lambda_{q',1} - j(\bar{\sigma}_{i,1} \cdot \hat{z})] \tag{22}$$

where δ_{i0} is the Kronecker delta function and the value of p' is chosen to correspond to the i-th wave. For the boundary between the n-th and n+1-th slab gratings (z = nd/N), the boundary condition for tangential E is

$$\sum_{q'=1}^{2s} C_{q',n}\, w_{p',q',n} \exp\{[\lambda_{q',n} - j(\bar{\sigma}_{i,n} \cdot \hat{z})]\, nd/N\} =$$

$$\sum_{q'=1}^{2s} C_{q',n+1}\, w_{p',q',n+1} \exp\{[\lambda_{q',n+1} - j(\bar{\sigma}_{i,n+1} \cdot \hat{z})] nd/N \qquad (23)$$

and for tangential H is

$$\sum_{q'=1}^{2s} C_{q',n}\, w_{p',q',n}\, [\lambda_{q',n} - j(\bar{\sigma}_{i,n} \cdot \hat{z})] \exp\{[\lambda_{q',n} - j(\bar{\sigma}_{i,n} \cdot \hat{z})]\, nd/N\} =$$

$$\sum_{q'=1}^{2s} C_{q',n+1}\, w_{p',q',n+1} [\lambda_{q',n+1} - j(\bar{\sigma}_{i,n+1} \cdot \hat{z})] \cdot \qquad (24)$$

$$\exp\{[\lambda_{q',n+1} - j(\bar{\sigma}_{i,n+1} \cdot \hat{z})]\, nd/N\}$$

For the boundary between the N-th slab grating and region 3 (z=d), the boundary condition for tangential E is

$$\sum_{q'=1}^{2s} C_{q',N}\, w_{p',q',N} \exp\{[\lambda_{q',N} - j(\bar{\sigma}_{i,N} \cdot \hat{z})]\, d\} = T_i \qquad (25)$$

and for tangential H is

$$\sum_{q'=1}^{2s} C_{q',N}\, w_{p',q',N}\, [\lambda_{q',N} - j(\bar{\sigma}_{i,N} \cdot \hat{z})] \exp\{[\lambda_{q',N} - j(\bar{\sigma}_{i,N} \cdot \hat{z})]\, d\} =$$

$$-j(\bar{k}_{3i} \cdot \hat{z})\, T_i \qquad (26)$$

Equations (21) through (26) represent a total of 2(N+1)s equations. There are s unknown values each of R_i and T_i and 2s unknown values of $C_{q',n}$ for

each slab grating. Thus, the total number of unkowns is 2(N+1)s, the same as the number of boundary condition equations. If s values of i are retained in the analysis, then the calculations will yield s transmitted wave amplitudes (T_i) and s reflected wave amplitudes (R_i).

An efficient procedure to solve this large system of equations is to use a technique like Gauss elimination[21] applied successively to each boundary starting at the z=0 input surface. By using this technique N+1 times in sequence, the s values of R_i and s values of T_i may be obtained in a single pass on the last step. As depicted in Fig. 3 the boundary condition equations are written as a matrix equation. The matrix is 2(N+1)s by 2(N+1)s and consists of the coefficients of $C_{q',n}$ (for q' = 1 to 2s and n = 1 to N), R_i (s values), and T_i (s values). For each slab grating, the boundary condition equations for its two boundaries produces a 4s by 2s submatrix. Starting with the first (n=1) slab grating (represented by upper left submatrix), a technique like Gauss elimination is applied to make all of the elements of the lower half of the submatrix equal to zero. This reduces the system to 2Ns equations. Repeating this procedure on the next (n=2) 4s by 2s submatrix reduces the system to 2(N-1)s equations. This process is continued until after N steps, only 2s equations in the diffracted amplitudes, R_i and T_i remain. These are then solved for R_i and T_i. At each step in this sequential process, a new set of coefficients of R_i are produced as shown by the dashed box in Fig. 3. After N steps these coefficients have moved to the bottom of the matrix and the final set of 2s equations in R_i and T_i are formed. This sequential procedure enormously reduces the storage and computational requirements for this type of problem. At each step, only a small 4s by 2s matrix is

being treated as opposed to the entire 2(N+1)s by 2(N+1)s matrix where N might typically be 50 to 100.

When the amplitudes R_i and T_i are known, then the diffraction efficiencies (ratio of diffracted intensity to input intensity) may be directly determined. The diffraction efficiencies in regions 1 and 3 are

$$DE_{1i} = \text{Re}\ [(\bar{k}_{1i} \cdot \hat{z})/(\bar{k}_{10} \cdot \hat{z})]\ R_i R_i^* \tag{27}$$

and

$$DE_{3i} = \text{Re}\ [(\bar{k}_{3i} \cdot \hat{z})/(\bar{k}_{10} \cdot \hat{z})]\ T_i T_i^* \tag{28}$$

The real part of the ratio of the propagation constants occurs when the time-average power-flow density is obtained by taking the real part of the complex Poynting vector. For lossless gratings the input power is conserved and thus the sum of all of the efficiencies for the propagating waves is unity. That is,

$$\sum_i (DE_{1i} + DE_{3i}) = 1 \quad . \tag{29}$$

This relationship may be used to verify the convergence of the numerical calculations.

4. RESULTS AND DISCUSSION

Using the method of solution described in the previous section, it is possible to calculate the diffraction efficiencies of dielectric surface relief gratings to an arbitrary level of accuracy. The analysis

contains no restrictions with respect to grating profile, groove depth, angle of incidence, or wavelength. Example results are presented in this section for sinusoidal (Fig. 4), square-wave (Fig. 5), triangular (Fig. 6), and sawtooth (Figs. 7 and 8) gratings. In each case the input wave is a plane wave with polarization perpendicular to the plane of incidence (H-mode) and incident at the first Bragg angle (m=1). For the cases presented, $\lambda=\Lambda$ and so the first Bragg angle is $\theta' = 30°$. Unlike previous methods, large groove depths do not cause numerical instabilities or inaccuracies. Results for groove depths as deep as four grating periods ($d/\Lambda = 4$) are presented in Figs. 4 to 8. Previous published results have had maximum groove depths of $d/\Lambda = 0.60$[5], 1.0[14], and 2.4.[16] The input region has a relative permittivity of $\varepsilon_I = 1.00$ (air) and the substrate a relative permittivity of $\varepsilon_{III} = 2.50$ (refractive index of 1.58).

With the exception of the zero-order wave, the reflected waves have diffraction efficiencies of less than 1%. The zero-order (specularly reflected) wave generally was found to decrease in intensity from the Fresnel reflection value with increasing groove depth. In the square-wave grating case, however, the zero-order reflected intensity is seen to oscillate with groove depth.

For the combination of relative permittivities used ($\varepsilon_I = 1.00$ and $\varepsilon_{III} = 2.50$), the transmitted first-order ($i = +1$) diffraction efficiency was found to be maximum for a groove depth in the range from 1.5 to 2 grating periods. For the sinusoidal, square-wave, and triangular grating profiles, the first-order diffraction efficiency was found to reach nearly 100% for the properly chosen groove depth. For the sawtooth gratings, however, the maximum first order diffraction efficiency was only about 50%. A result that was consistently found in this work is: Dielectric

gratings with profiles that are expressable as even functions are capable of high (greater than 85%) diffraction efficiency. Thus, if the grating profile is symmetric so that the x=0 origin can be chosen so that F(x) is an even function, then large first-order diffraction efficiencies are possible. The diffraction efficiencies calculated for any case are, of course, independent of the choice of the x=0 origin. But if the grating has even symmetry, large diffraction efficiencies are possible. For grating profiles that lack even symmetry, such as the sawtooth profile, the maximum diffraction efficiencies are correspondingly less. Maximum diffraction efficiency results for several profiles are given in Table I. Notice that the stairstep grating profile of even symmetry has maximum first-order diffraction efficiency of 89.4%. However, the same stairstep grating with steps shifted producing a profile of only odd symmetry has a maximum first-order diffraction efficiency of only 67.7% or 71.8% (depending on orientation of Bragg angle). The lack of even symmetry obviously produces dephasing of the fundamental first-order diffracted beam and causes the diffraction efficiency not to reach a large value. Because the lack of even symmetry is less severe in the stairstep than in the sawtooth, the maximum efficiency for the stairstep is correspondingly larger. Further, it is noted from Table I that the choice of orientation of an asymmetric grating with respect to the Bragg angle does not appreciably affect the maximum possible diffraction efficiency. There is a modest trend among the first four gratings in Table I (each expressable as an even function) that indicates higher maximum diffraction efficiencies are obtainable with sharper, more pointed grating profiles. Thus in the sequence of square-wave, stairstep, sinusoid, and triangular grating, the maximum diffraction efficiency steadily increases (from 88.5% to 99.0%). Again,

however, the dominant factor controlling the maximum diffraction efficiency is the degree of even symmetry in the grating profile.

The method of analysis presented in this paper is rigorous and any arbitrary level of accuracy is attainable if needed. The method can be applied to any grating profile. Further, it is straightforward to add substrate and cover layers using this method of analysis if they are present.

5. NUMERICAL ACCURACY

In this section the accuracy, the convergence, and the stability of the numerical technique are discussed. Comparisons to previously published results from other techniques are also presented.

For all of the numerical results presented, both the number of waves (space harmonics) included and the number of layers were increased until the changes in the calculated diffraction efficiencies were less than 10^{-8}. Typically 50 to 100 layers and 8 to 12 waves (values of i) were used. The algorithm described was <u>always</u> stable and it converged in <u>every</u> case considered. The conservation of power was also always accurately achieved. In all of our results the error in the conservation of power, Eq. (29), was on the order of 10^{-12}.

Although the resulting system of simultaneous linear equations is 2(N+1)s by 2(N+1)s, the technique described in this paper handles this very large system by reducing it to the problem of converting a 4s by 2s system into upper matrix form N times. This reduces the data storage requirement by a factor of $(N+1)^2/3$ and reduces the execution time by a factor of N^2 where N is the number of layers. Without such savings, it would be extremely difficult to solve this large system of equations on even the

largest of computers. Using the present method, a typical execution time for an 8-wave, 50-layer configuration is about 6 seconds for each diffraction efficiency point on a CDC 730/760 computer.

In the present work the rigorous coupled-wave approach is used to find the fields in each of the layers of the grating. Another possible technique is that of Lewis and Hessel[18] which analyzes a periodic array of dielectric slabs. The coupled-wave approach is as accurate and as rigorous as the array of slabs technique. Both require finding the eigenvalues of a truncated characteristic or coupled-wave equation matrix. However, the coupled-wave approach is more suitable for computer implementation.

To show the accuracy and the effectiveness of the present technique, results obtained in the present work were compared to those results that are available in the literature. Our results show excellent agreement with the graphical results of Tremain and Mei[14] (their Figs. 8-10). Of course, this agreement is limited by the accuracy of the graphs. In Table II, results calculated by the present method are compared to numerical results from Chang, Shah, and Tamir[16] (their Table I). As can be seen, our results agree extremely well with theirs. However, in the present work the deviation from the conservation of power ($\sim 10^{-12}$) is much smaller than either that of van den Berg[13] ($\sim 10^{-4}$) or Chang, Shah, and Tamir[16] ($\sim 10^{-6}$). In Table III results obtained by the present rigorous coupled-wave approach are compared to those presented by Tremain and Mei[14] (their Table II). It is shown that the deviations from the conservation of power in the present results ($\sim 10^{-12}$) are much smaller than those of Zaki[12] ($\sim 10^{-2}$) or of Tremain and Mei[14] ($\sim 10^{-3}$). The small size of the residual error in the conservation of power in the present work provides additional strong evidence that the diffraction efficiency results are similarly accurate.

Thus it may be concluded that the rigorous coupled-wave method of analysis of surface-relief gratings is accurate, stable, reliable, and efficient.

This work was sponsored in part by the National Science Foundation and the Joint Services Electronics Program.

REFERENCES

1. R. Petit (ed.), Electromagnetic Theory of Gratings, (Springer-Verlag, Berlin, 1980).
2. M. Neviere, R. Petit, and M. Cadilhac, "About the theory of optical grating coupler-waveguide systems," Opt. Commun. 8, 113-117 (1973).
3. S. T. Peng, H. L. Bertoni, and T. Tamir, "Analysis of periodic thin-film structures with rectangular profiles," Opt. Commun. 10, 91-94 (1974).
4. S. T. Peng and T. Tamir, "Directional blazing of waves guided by asymmetrical dielectric gratings," Opt. Commun. 11, 405-409 (1974).
5. S. T. Peng, T. Tamir, and H. L. Bertoni, "Theory of periodic dielectric waveguides," IEEE Trans. Microwave Theory Tech. MTT-23, 123-133 (1975).
6. D. Marcuse, "Exact theory of TE-wave scattering from blazed dielectric gratings," Bell Syst. Tech. J. 55, 1295-1317 (1976).
7. K. Knop, "Rigorous diffraction theory for transmission phase gratings with deep rectangular grooves," J. Opt. Soc. Am. 68, 1206-1210 (1978).
8. K. C. Chang and T. Tamir, "Simplified approach to surface-wave scattering by blazed dielectric gratings," Appl. Opt. 19, 282-288 (1980).
9. W. Streifer, D. R. Scifres, and R. D. Burnham, "Analysis of grating-coupled radiation in GaAs:GaAlAs lasers and waveguides," IEEE J. Quantum Electron. QE-12, 422-428 (1976).
10. T. Tamir and S. T. Peng, "Analysis and design of grating couplers," Appl. Phys. 14, 235-254 (1977).

11. A. R. Neureuther and K. Zaki, "Numerical methods for the analysis of scattering from nonplanar and periodic structures," Alta Freq. 38, 282-285 (1969).

12. K. Zaki, "Numerical methods for the analysis of scattering from nonplanar periodic structures," Ph.D. thesis, University of California Berkeley (1969) (unpublished).

13. P. M. van den Berg, "Rigorous diffraction theory of optical reflection and transmission gratings," Ph.D. thesis, Delft University of Technology Netherlands, Report No. 1971-16 (1971) (unpublished).

14. D. E. Tremain and K. K. Mei, "Application of the unimoment method to scattering from periodic dielectric structures," J. Opt. Soc. Am. 68, 775-783 (1978).

15. K. C. Chang, "Surface-wave scattering by dielectric gratings with arbitrary profiles," Ph.D. thesis, Polytechnic Institute of New York, Brooklyn (1979) (unpublished).

16. K. C. Chang, V. Shah, and T. Tamir, "Scattering and guiding of waves by dielectric gratings with arbitrary profiles," J. Opt. Soc. Am. 70, 804-813 (1980).

17. M. G. Moharam and T. K. Gaylord, "Rigorous coupled-wave analysis of planar grating diffraction," J. Opt. Soc. Am. 71, 811-818 (1981).

18. L. R. Lewis and A. Hessel, "Propagation characteristics of periodic arrays of dielectric slabs," IEEE Trans. Microwave Theory Tech. MTT-19, 276-286 (1971).

19. e.g. C. L. Liu and J. W. S. Liu, Linear Systems Analysis (McGraw-Hill, New York, 1975).

20. e.g. program EIGRF from the International Mathematics and Statistics Library (IMSL).

21. e.g. B. Carnahan, H. A. Luther, and J. O. Wilkes, Applied Numerical Methods (Wiley, New York, 1969).

FIGURE CAPTIONS

Fig. 1. Geometry of dielectric surface-relief grating.

Fig. 2. The n-th planar grating resulting from the decomposition of the surface-relief grating into N thin gratings.

Fig. 3. Matrix equation representation of 2(N+1)s boundary condition equations where s is the total number of diffracted waves retained in the analysis. C_1 represents the column vector $C_{q',1}$ where $q' = 1$ to 2s and likewise for C_2 through C_N. The reflected and transmitted amplitudes, R_i and T_i, are column vectors of length s. The product output vector, before manipulation, is all zeros except for the two ones that are shown. These correspond to the normalized E and H values in the input wave.

Fig. 4. Diffraction efficiency as a function of groove depth for a lossless sinusoidal surface relief grating. Incidence is at first Bragg angle (m=1), $\lambda = \Lambda$, $\varepsilon_I = 1.00$, and $\varepsilon_{III} = 2.50$.

Fig. 5. Diffraction efficiency as a function of groove depth for a lossless square-wave surface-relief grating. Incidence is at first Bragg angle (m=1), $\lambda = \Lambda$, $\varepsilon_I = 1.00$, and $\varepsilon_{III} = 2.50$.

Fig. 6. Diffraction efficiency as a function of groove depth for a lossless triangular grating. Incidence is at first Bragg angle (m=1), $\lambda = \Lambda$, $\varepsilon_I = 1.00$, and $\varepsilon_{III} = 2.50$.

Fig. 7. Diffraction efficiency as a function of groove depth for a lossless sawtooth grating. Incidence is at first Bragg angle (m=1) with the inclined surfaces of the sawtooth facing the incident beam. Also $\lambda = \Lambda$, $\varepsilon_I = 1.00$, and $\varepsilon_{III} = 2.50$.

Fig. 8. Diffraction efficiency as a function of groove depth for a lossless sawtooth grating. Incidence is at first Bragg angle (m=1) with the inclined surfaces of the sawtooth facing away from the incident beam. Also $\lambda = \Lambda$, $\varepsilon_I = 1.00$, and $\varepsilon_{III} = 2.50$.

Table I. Maximum transmitted first-order (i = +1) diffraction efficiencies for various grating profiles. Incidence is at first Bragg angle, $\theta' = 30°$, $\varepsilon_I = 1.00$, and $\varepsilon_{III} = 2.50$. Diffraction efficiencies correspond to the value at the first peak of the $DE_{3,1}$ versus d/Λ curve.

	$(DE_{3,1})_{max}$	$\frac{d}{\Lambda}$
	99.0	2.10
	95.9	1.75
	89.4	1.60
	88.5	1.55
	71.8	1.62
	67.7	1.67
	51.0	2.10
	50.6	2.10

Table II. Comparison of diffraction efficiencies calculated by van den Berg[13] and Chang, Shah, Tamir[16] to values calculated in the present work. The sinusoidal surface-relief grating is characterized by $\varepsilon_I = 1.0$ and $\varepsilon_{III} = 4.0$. The incident plane wave is characterized by $\lambda = \Lambda$ and $\theta' = 45°$.

Diffraction Efficiency	d/Λ	van den Berg	Chang, Shah, Tamir	Present Work
$DE_{1,0}$	0.005	0.203990	0.2037150	0.2037170
$DE_{1,1}$	0.005	0.000027	0.0000265	0.0000262
$DE_{3,-1}$	0.005	0.000074	0.0000817	0.0000820
$DE_{3,0}$	0.005	0.795207	0.7961250	0.7961205
$DE_{3,1}$	0.005	0.000056	0.0000550	0.0000543
$DE_{3,2}$	0.005	0.000000	1.3×10^{-9}	1.3×10^{-9}
$\sum DE-1$	0.005	-1.1×10^{-4}	3.6×10^{-6}	3.3×10^{-12}
$DE_{1,0}$	0.5	0.02418	0.0349913	0.0346051
$DE_{1,1}$	0.5	0.06367	0.0537149	0.0537065
$DE_{3,-1}$	0.5	0.11222	0.1693248	0.1690036
$DE_{3,0}$	0.5	0.39340	0.3492882	0.3484211
$DE_{3,1}$	0.5	0.34700	0.3307839	0.3323272
$DE_{3,2}$	0.5	0.05563	0.0618147	0.0619366
$\sum DE-1$	0.5	-3.9×10^{-3}	8.2×10^{-5}	1.7×10^{-12}

Table III. Comparison of diffraction efficiencies calculated by Tremain and Mei[14] and Zaki[12] to values calculated in the present work. The sinusoidal surface-relief grating is characterized by $\varepsilon_I = 1.0$, $\varepsilon_{III} = 2.3104$, and $d/\Lambda = 1.1765$. The plane wave is incident at $\theta' = 36°$. The sum of the diffraction efficiencies includes other propagating waves even though these are not listed separately in this table.

Diffraction Efficiency	λ/Λ	Tremain, Mei	Zaki	Present Work
$DE_{3,0}$	0.750	0.202	0.16	0.15377
$DE_{3,1}$	0.750	0.547	0.50	0.57732
$DE_{3,2}$	0.750	0.241	0.29	0.20013
$\sum$DE-1	0.750	7×10^{-3}	-4×10^{-2}	1.35×10^{-12}
$DE_{3,0}$	0.941	0.301	0.30	0.30290
$DE_{3,1}$	0.941	0.647	0.62	0.63689
$DE_{3,2}$	0.941	0.050	0.03	0.05058
$\sum$DE-1	0.941	8×10^{-3}	-4×10^{-2}	1.34×10^{-12}
$DE_{3,0}$	1.030	0.339	0.33	0.34202
$DE_{3,1}$	1.030	0.643	0.555	0.62361
$DE_{3,2}$	1.030	0.023	0.01	0.02398
$\sum$DE-1	1.030	1×10^{-2}	9×10^{-2}	8.6×10^{-13}
$DE_{3,0}$	1.177	0.396	0.39	0.39857
$DE_{3,1}$	1.177	0.598	0.58	0.59051
$DE_{3,2}$	1.177	0.000	0.00	0.00000
$\sum$DE-1	1.177	6×10^{-3}	2×10^{-2}	1.36×10^{-12}
$DE_{3,0}$	1.471	0.493	0.49	0.49355
$DE_{3,1}$	1.471	0.497	0.47	0.49134
$DE_{3,2}$	1.471	0.000	0.00	0.00000
$\sum$DE-1	1.471	5×10^{-3}	2×10^{-2}	1.03×10^{-12}

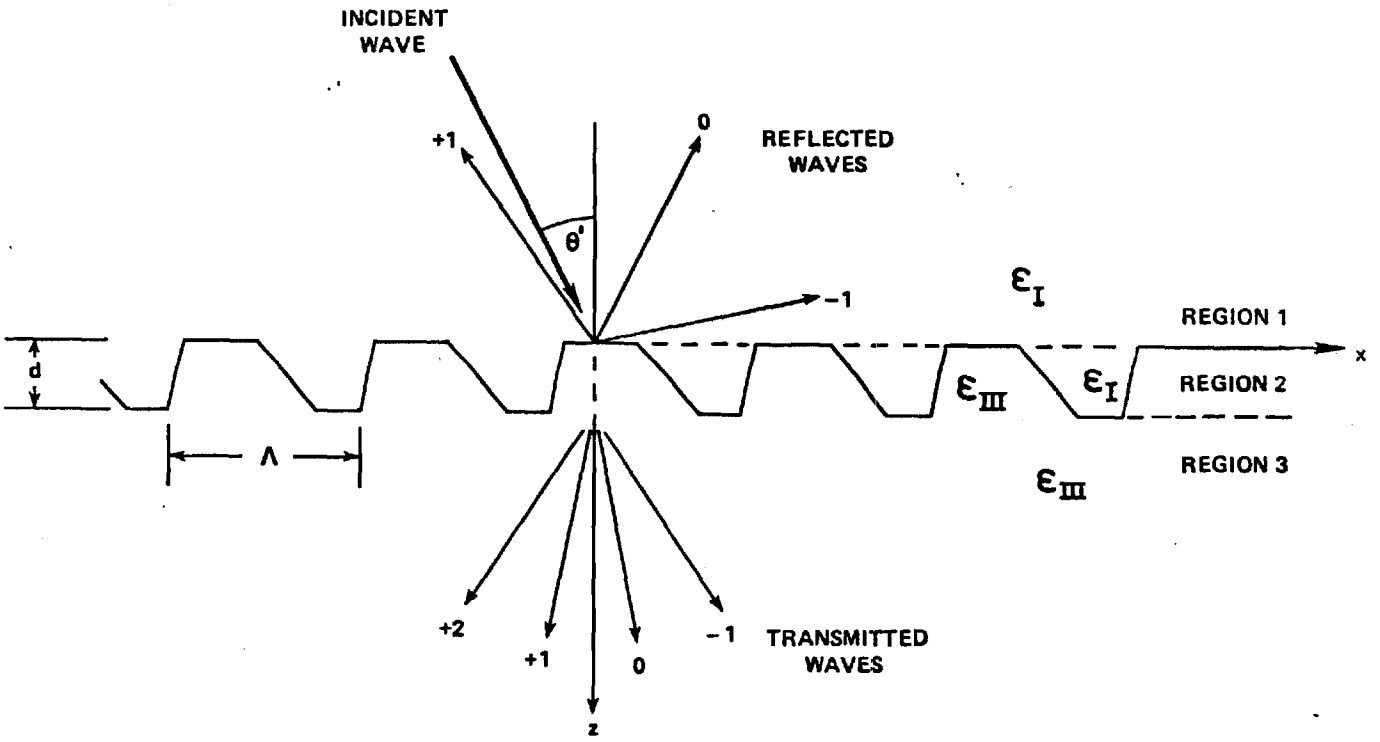

INCIDENT
WAVE
+1
0
REFLECTED
WAVES
θ'
-1
ε_I
REGION 1
x
d
ε_{III}
ε_I
REGION 2
Λ
REGION 3
ε_{III}
+2
+1
0
-1
TRANSMITTED
WAVES
z

Fig.1

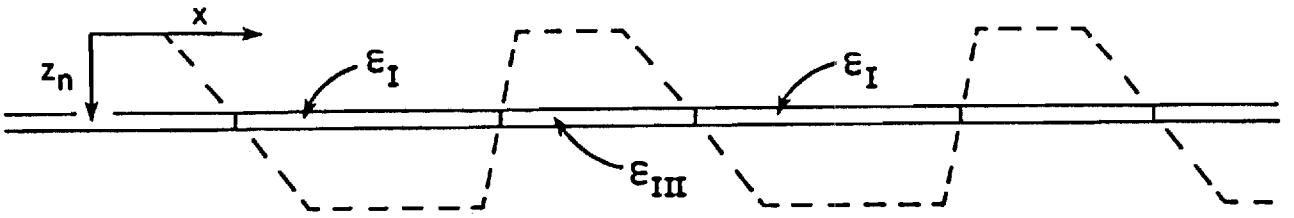


Fig.2

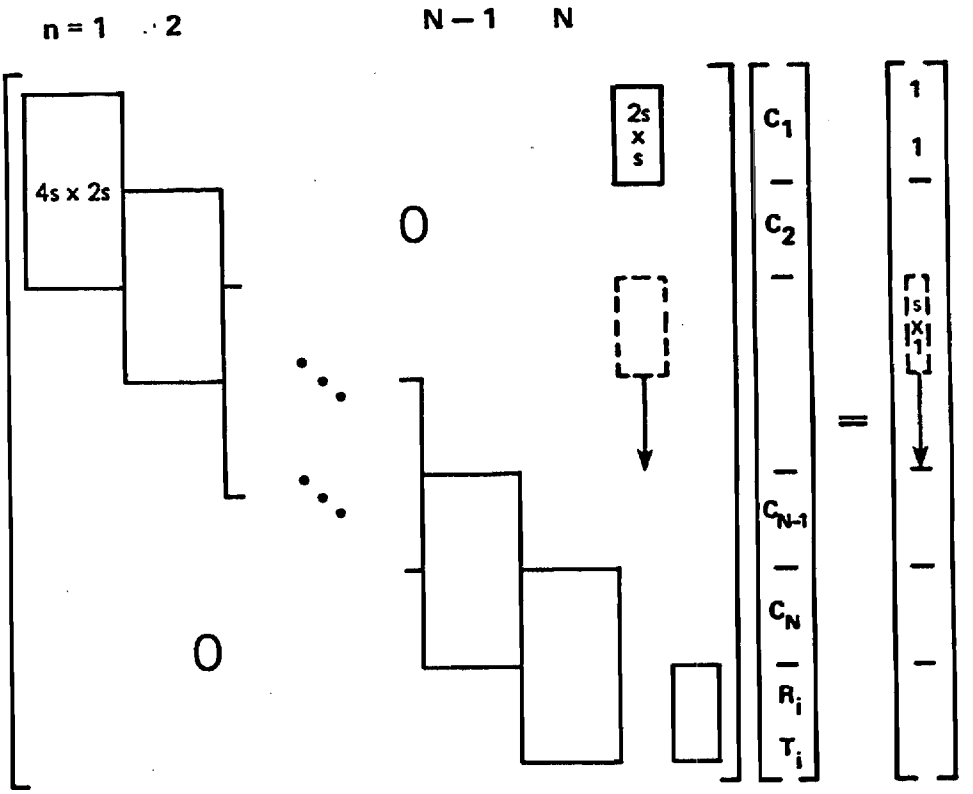


Fig.3

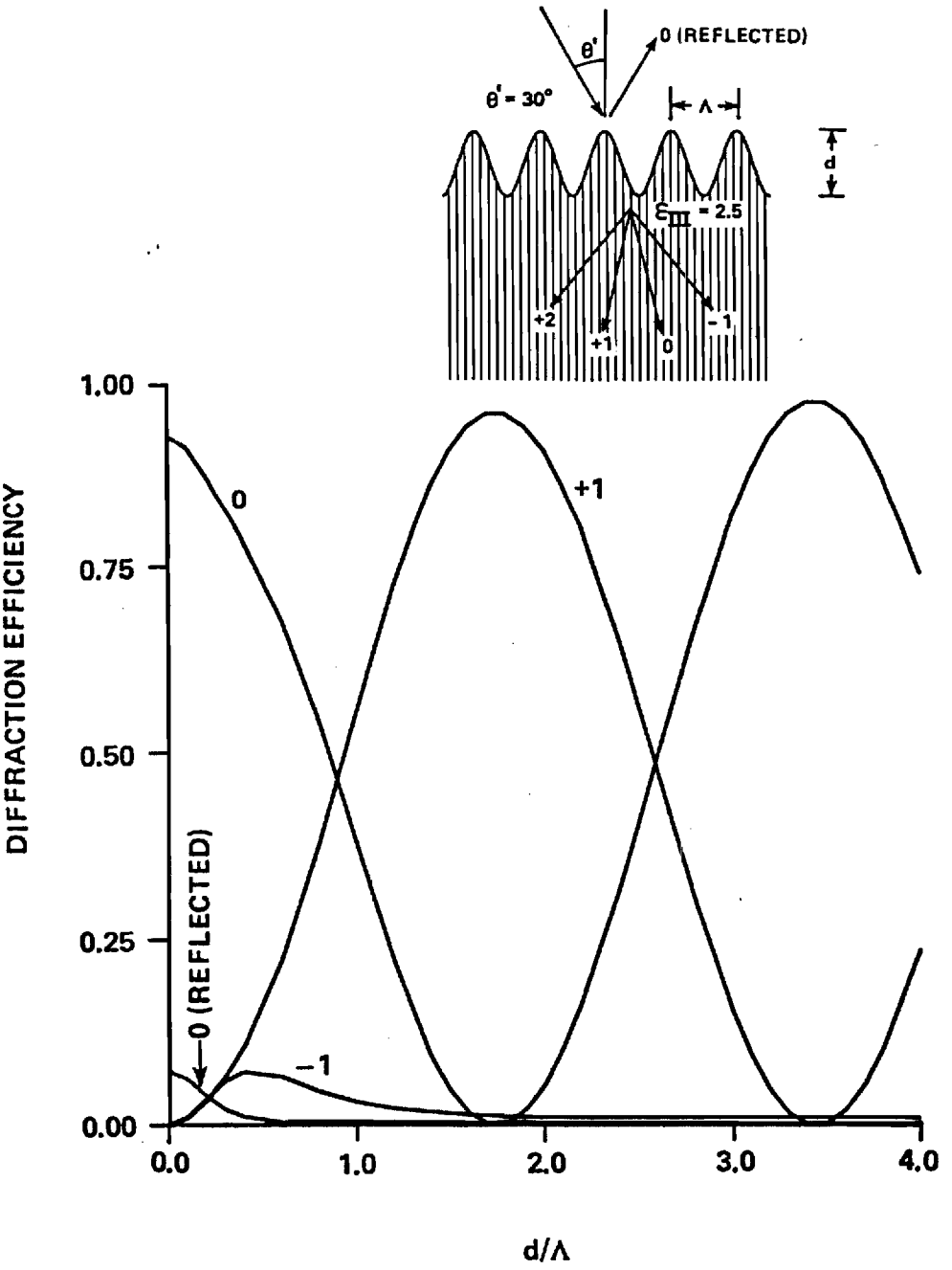

θ'
0 (REFLECTED)
θ' = 30°
Λ
d
ε_{III} = 2.5
+2
+1
0
-1
1.00
0.75
0.50
0.25
0.00
DIFFRACTION EFFICIENCY
0
+1
0 (REFLECTED)
-1
0.0
1.0
2.0
3.0
4.0
d/Λ

Fig.4

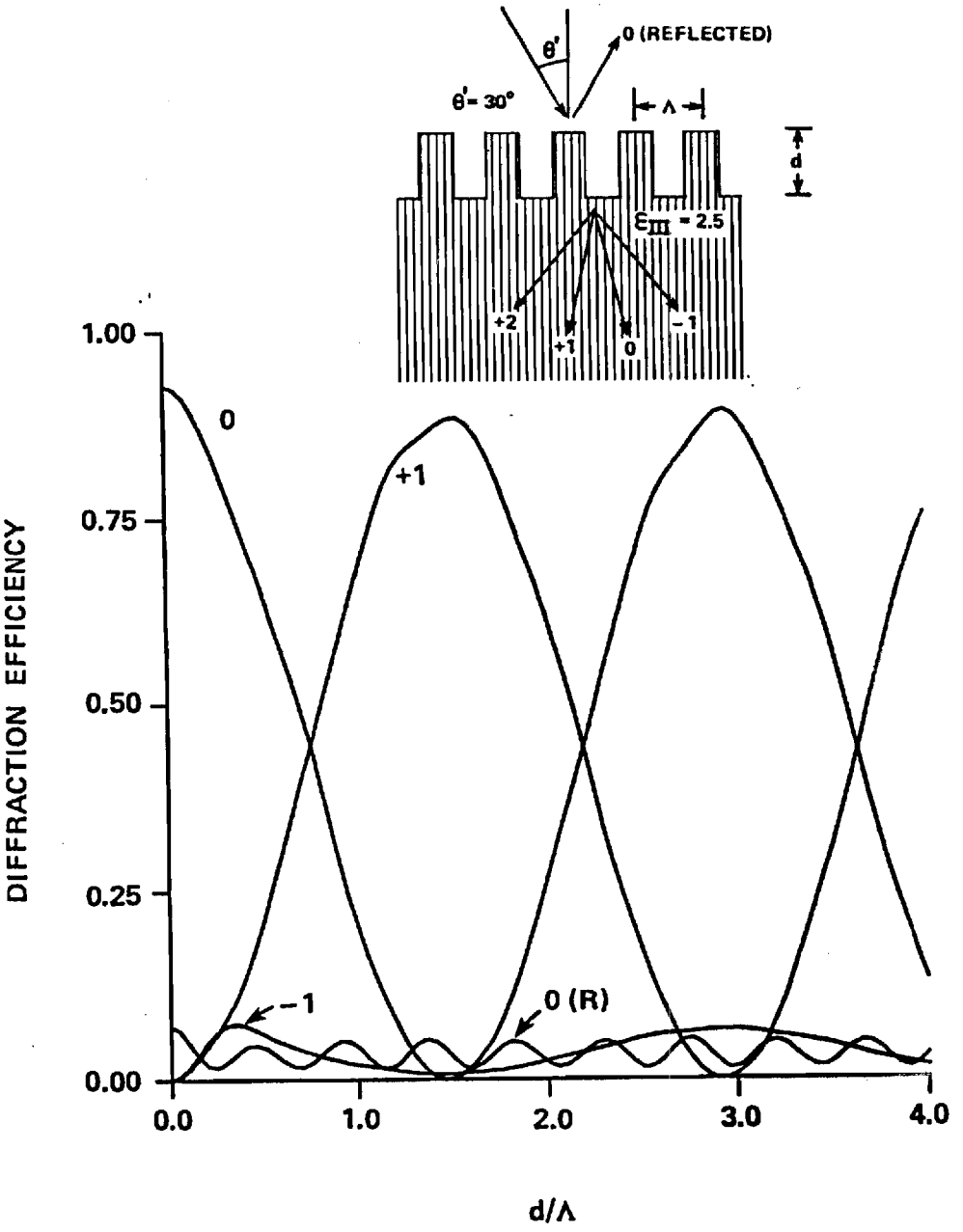

θ'
0 (REFLECTED)
θ'= 30°
Λ
d
ε_{III} = 2.5
+2
+1
0
−1
1.00
0.75
0.50
0.25
0.00
0
+1
−1
0 (R)
DIFFRACTION EFFICIENCY
0.0
1.0
2.0
3.0
4.0
d/Λ

Fig.5

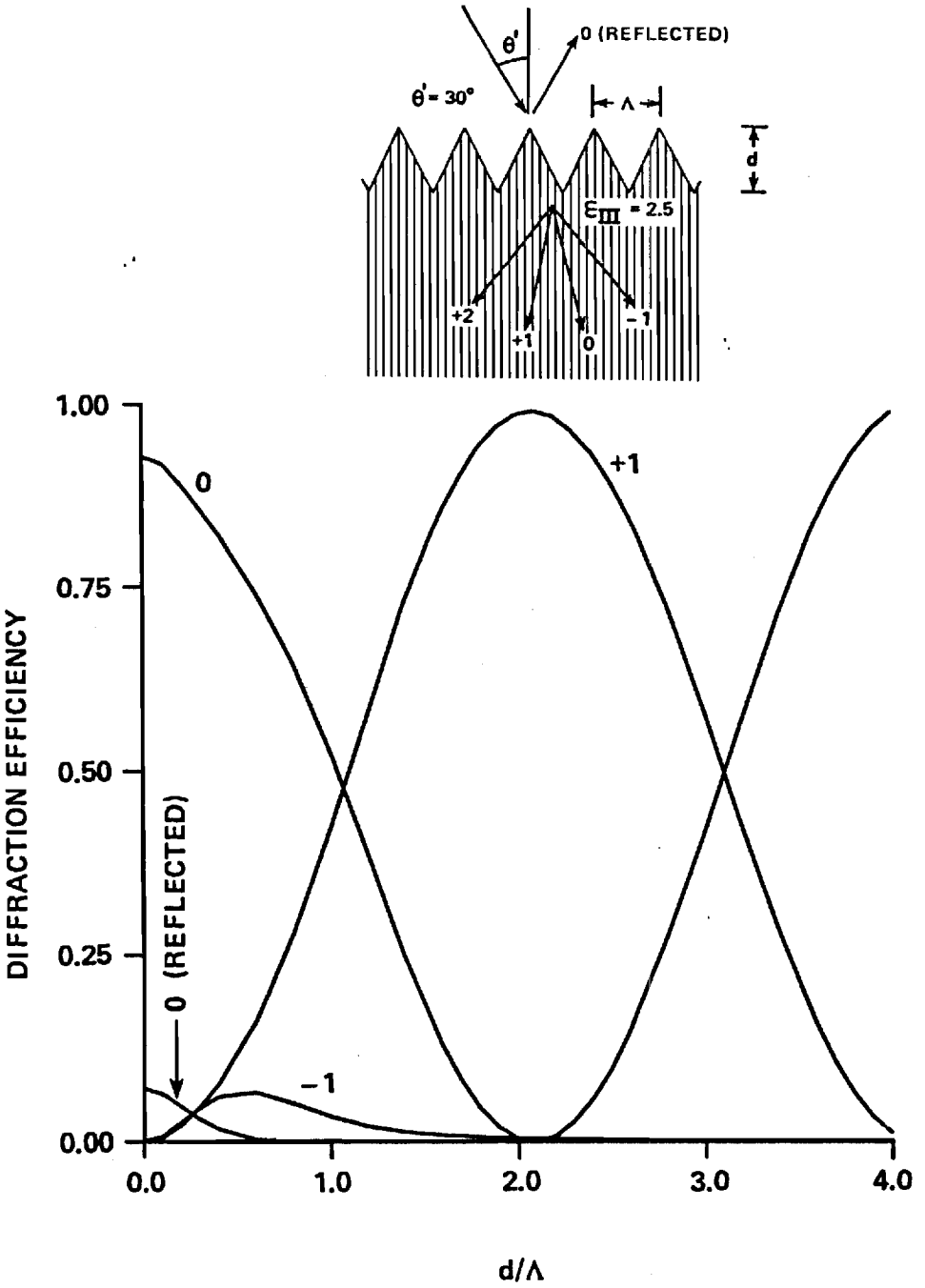

θ'
0 (REFLECTED)
θ' = 30°
Λ
d
ε III = 2.5
+2
+1
0
−1
1.00
0.75
0.50
0.25
0.00
DIFFRACTION EFFICIENCY
0
+1
0 (REFLECTED)
−1
0.0
1.0
2.0
3.0
4.0
d/Λ

Fig.6

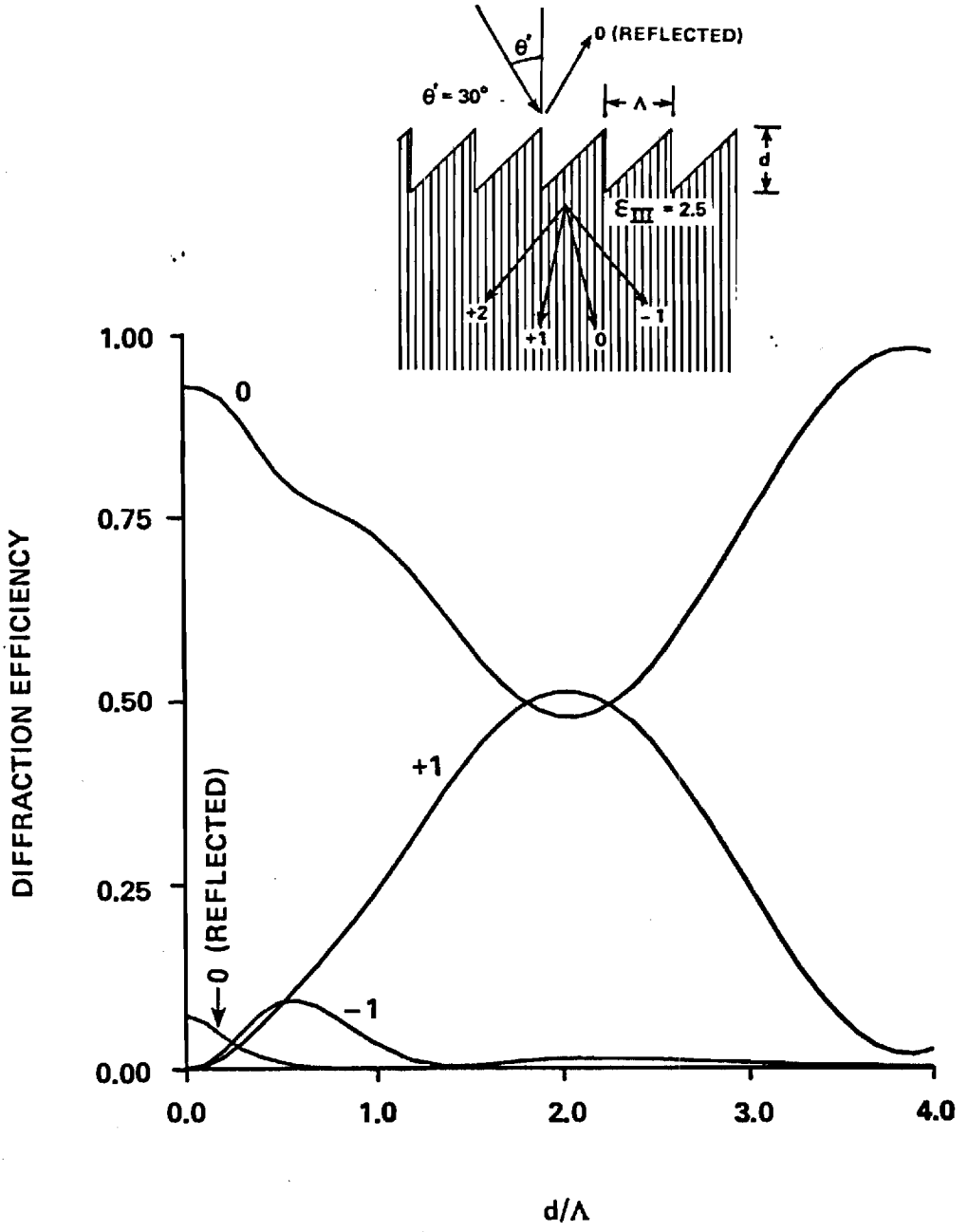

θ'
0 (REFLECTED)
θ' = 30°
Λ
d
ε III = 2.5
+2
+1
0
-1
1.00
0.75
0.50
0.25
0.00
0
+1
-1
0 (REFLECTED)
DIFFRACTION EFFICIENCY
0.0
1.0
2.0
3.0
4.0
d/Λ

Fig.7

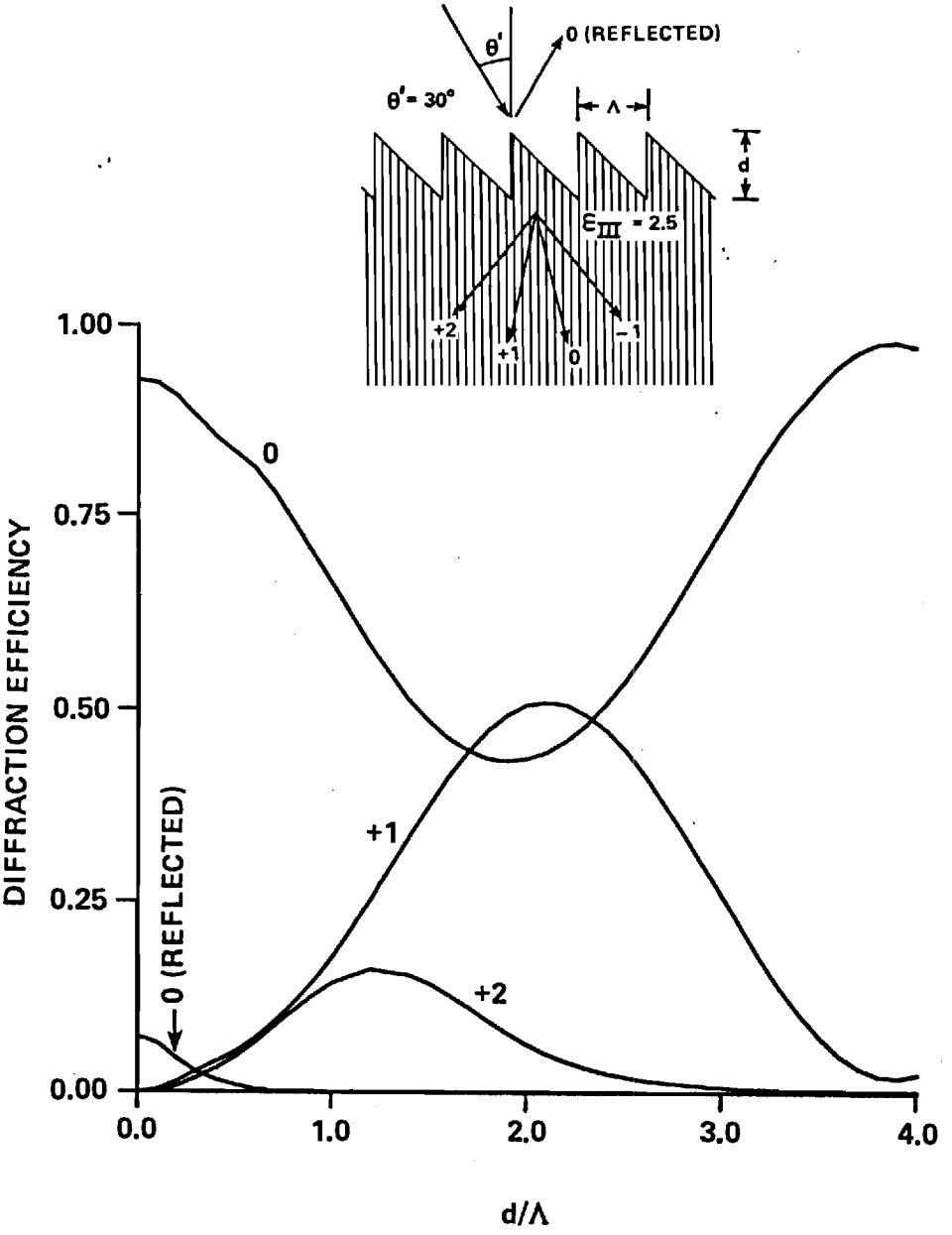

θ'
0 (REFLECTED)
θ' = 30°
Λ
d
ε_{III} = 2.5
+2
+1
0
−1
1.00
0.75
0.50
0.25
0.00
DIFFRACTION EFFICIENCY
0
+1
+2
0 (REFLECTED)
0.0
1.0
2.0
3.0
4.0
d/Λ

Fig.8

Appl. Phys. B 28, 1–14 (1982)

Applied Physics B Photophysics and Laser Chemistry

Planar Dielectric Grating Diffraction Theories

T. K. Gaylord and M. G. Moharam

School of Electrical Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA

Received 23 December 1981/Accepted 26 February 1982

Abstract. Various planar dielectric grating diffraction theories are reviewed for the case of a general sinusoidal permittivity planar grating with slanted fringes and plane wave incidence at an arbitrary angle. Exact formulations without approximations (rigorous coupled-wave analysis and rigorous modal analysis) are developed first. Then, using a series of fundamental assumptions, rigorous theory is shown to reduce to the various approximate theories in the appropriate limits. The implications of these fundamental assumptions are discussed.

PACS: 42.10, 42.20, 42.30

Since 1930 there have been over 400 scientific papers on the subject of grating diffraction. Many of these papers have been applicable to planar dielectric gratings. These periodic structures have been applied in numerous areas such as acousto-optics, holography, integrated optics, and spectral analysis. The diffraction of electromagnetic waves by spatially periodic media may be analyzed by numerous methods and with a wide variety of possible assumptions. The purpose of this paper is to review both rigorous and approximate planar grating diffraction theories and to show explicitly the relationships between the various theories.

The most common methods of analyzing planar dielectric grating diffraction are the coupled-wave approach [1–8] and the modal approach [9–18]. These theories have recently been treated in two extensive reviews [19, 20]. Both coupled-wave and modal approaches can produce exact formulations without approximations. In their full rigorous forms these formulations are completely equivalent [21]. They represent merely alternative methods of representing the electromagnetic fields inside the grating (Sect. 2).

Starting with the rigorous theories and using a series of fundamental assumptions, these general theories are shown to reduce to the various approximate theories [two-wave modal theory, two-wave second-order coupled-wave theory, multiwave coupled-wave theory, two-wave first-order coupled-wave theory (Kogelnik theory), Raman-Nath theory, and amplitude transmittance theory] in the appropriate limits. This is shown in Sect. 8.

1. Planar Dielectric Grating Diffraction

The general planar grating diffraction problem is depicted in Fig. 1. An electromagnetic wave is obliquely incident upon a slanted-fringe planar grating bounded by two different homogeneous media. In general, there will be simultaneously both forward-diffracted and backward-diffracted waves as shown in the figure. This geometry is applicable 1) to holographic gratings in air or other media ($\varepsilon_I = \varepsilon_{III} \neq \varepsilon_0$), 2) to acousto-optic gratings within a medium ($\varepsilon_I = \varepsilon_0 = \varepsilon_{III}$), and 3) to grating couplers such as used in integrated optics ($\varepsilon_I \neq \varepsilon_0 \neq \varepsilon_{III} \neq \varepsilon_I$). The quantities ε_I, ε_0, and ε_{III} are the average relative permittivities (dielectric constants) in regions 1, 2, and 3, respectively.

In this paper, for simplicity, the case of a lossless dielectric grating with sinusoidal permittivity is treated. The incident plane wave polarization is perpendicular to the plane of incidence (H mode). This is probably the case of widest general interest. However, these assumptions are not essential to the theories described. The relative permittivity (dielectric constant) in the

0721-7269/82/0028/0001/$02.80

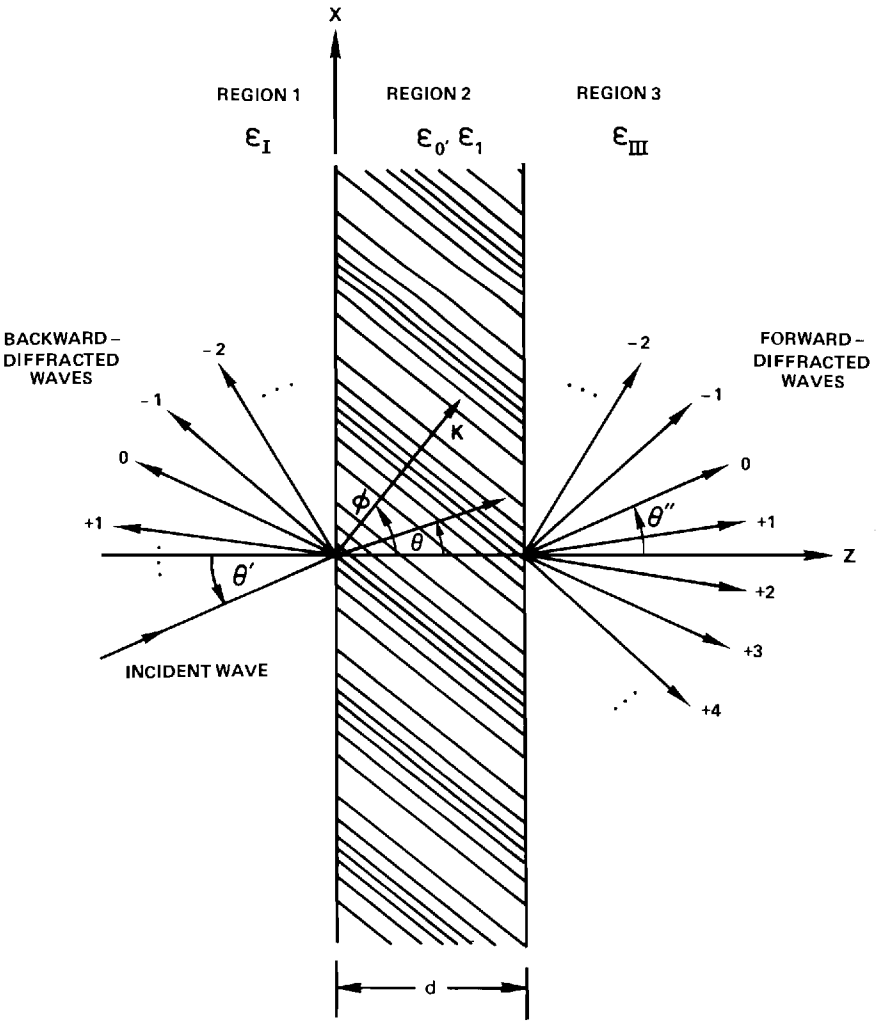


Fig. 1. Geometrical configuration of planar grating diffraction

grating region is given by

$$\varepsilon(x,z)=\varepsilon_0+\varepsilon_1\cos(\bar{K}\cdot\bar{r})$$
$$=\varepsilon_0+\varepsilon_1\cos[K(x\sin\phi+z\cos\phi)], \quad (1)$$

where ε_1 is the amplitude of the sinusoidal relative permittivity, ϕ is the grating slant angle, and $K=2\pi/\Lambda$, where Λ is the grating period. The cosinusoidal form used in (1) is common in the volume holographic grating literature. In the acousto-optics literature, a sinusoidal form for (1) is more common. Using the sinusoidal form would alter the resulting equations in the following sections as well as their amplitude solutions. However, the diffracted intensities are identical in either case.

The general approach to the planar-grating diffraction problem involves finding a solution of the wave equation in each of the three regions and then matching the tangential electric and magnetic fields at the two interfaces ($z=0$ and $z=d$). In region 1, the normalized amplitude of the incident plane wave is

$$E_{inc}=\exp(-j\bar{k}_1\cdot\bar{r})$$
$$=\exp[-jk_1(x\sin\theta'+z\cos\theta')], \quad (2)$$

where θ' is the angle of incidence on region 1, $k_1=2\pi(\varepsilon_1)^{1/2}/\lambda$, and λ is the free space wavelength. The wave equation for the H mode polarization is the scalar wave equation (Helmholtz equation)

$$\nabla^2E+k^2\varepsilon(x,z)E=0, \quad (3)$$

where $k=2\pi/\lambda$. For H mode polarization, the electric field only has a component in the y direction. The fields and the grating are unchanging in the y direction. For any arbitrary direction the grating as bounded by regions 1 and 3 is periodic only in the x component of the direction. If region 2 was infinite in all directions (not bounded) the resulting grating would be periodic in any direction that was not perpendicular to the grating vector $\bar{K}$. The Floquet theorem [22, 23] restricts the possible fields that can exist in a periodic structure at steady state. As a result of the Floquet theorem the diffracted wavevectors inside the grating may be represented for the infinite periodic medium case by

$$\bar{\sigma}_i=\bar{k}_2-i\bar{K}, \quad (4)$$

where $\bar{\sigma}_i$ is the wavevector of the i-th space harmonic in the grating, i is any integer, and $\bar{k}_2$ is the wavevector of the zero-order ($i=0$) space harmonic having a magnitude of $k_2=2\pi(\varepsilon_0)^{1/2}/\lambda$. For the bounded grating that is periodic only in the x-component of direction, the Floquet theorem only requires

$$\bar{\sigma}_i\cdot\hat{x}=(\bar{k}_2-i\bar{K})\cdot\hat{x}, \tag{5}$$

where $\hat{}$ denotes a unit vector. This expression is just the x-component of (4). Only the form as represented by (5) is necessary for the present problem. However, for subtle reasons that will become clear in the next section, the Floquet requirement as expressed by (4) will be used. This is certainly acceptable since it contains the necessary (5) within it.

2. Rigorous Coupled-Wave and Modal Theories

It is possible to formulate the planar grating diffraction problem depicted in Fig. 1 in an exact manner. This may be done with the coupled-wave approach or the modal approach. The modal approach is sometimes referred to as the Floquet, Floquet-Bloch, eigenmode, characteristic-mode, or coupled-mode approach. The coupled-wave approach is confusingly also sometimes called coupled-mode approach. Both the coupled-wave and modal approaches are altenrative methods of representing the fields inside the grating medium.

In the coupled-wave representation, the fields inside the grating are expanded in terms of the space harmonics of the fields in the periodic structure. These space harmonics inside the grating correspond to diffracted orders outside of the grating. Thus, the partial fields inside the modulated medium are visualized as diffracted waves that progress through the planar slab and couple energy back and forth between each other as they progress. This picture agrees rather well with simple physical intuition about the process of diffraction by a volume grating. In the coupled-wave approach the total field is thus expressed as

$$E(x,z)=\sum_{i=-\infty}^{+\infty} S_i(z)\exp(-j\bar{\sigma}_i\cdot\bar{r}), \tag{6}$$

where i is the space harmonic index. Equation (6) has the general appearance of a plane wave expansion of diffracted waves with amplitudes S_i. This would be true if the S_i's were constants. However, since the S_i's are not constants but are functions of z, each i does not correspond to a single plane wave. In general there are an infinite number of plane waves associated with each i. S_i varies only in the direction perpendicular to the boundary. The sum of all of the i-th partial fields as represented by $E(x,z)$ in (6) satisfies the wave equation. However, individually the partial fields do not satisfy the wave equation.

In the modal representation, the fields inside the grating are expanded in terms of the allowable modes of the periodic medium. The fields are visualized as waveguide modes in the grating region. In the modal approach, the total electric field is expressed as a weighted summation over all possible modes,

$$E(x,z)=\sum_{m=-\infty}^{+\infty} C_m\Phi_m(\bar{r})\exp(-j\bar{k}_{2m}\cdot\bar{r}), \tag{7}$$

where m is the mode index. The function $\Phi_m(\bar{r})$ is periodic with a period equal to the grating period. That is $\Phi_m(\bar{r})=\Phi_m(\bar{r}+\bar{\Lambda})$. The summation includes both forward and backward propagating modes. The backward propagating modes are due to diffraction in the grating volume (when the grating fringes are slanted) and due to reflections at the $z=d$ boundary. Each individual m-th mode satisfies the wave equation and may be either evanescent or propagating. These modes in the grating are precisely analogous to modes in a waveguide. Each waveguide mode satisfies the wave equation by itself and it may be either cutoff or propagating. Each mode (m) consists of an infinite number of space harmonics (i) and each mode propagates through the medium without change. The space harmonics may be viewed as arising from the Fourier expansion of the periodic function $\Phi_m(\bar{r})$.

The coupled-wave representation or expansion in terms of space harmonics, (6), and the modal representation or expansion in terms of modes, (7), are merely alternative representations of the same physical problem. Both approaches are complete and both are rigorous formulations (without approximations). These two approaches are thus completely equivalent and this will be demonstrated in Sect. 5.

3. Rigorous Coupled-Wave Equations

The rigorous coupled-wave equations are developed by starting with the field expansion in terms of space harmonics, (6). There are mutliple versions of the rigorous coupled-wave equations for the physical situation depicted in Fig. 1 depending on the form of the modulation chosen (sinusoidal or cosinusoidal permittivity), the form of the Floquet condition chosen, (4) or (5), and the polarization chosen (H mode or E mode). The basic case treated in this paper is cosinusoidal permittivity and H mode polarization. Both forms of the Floquet condition are analyzed in this section.

The Floquet condition for an infinite periodic medium, (4), will be treated first since this gives rise to the very useful, common form of the rigorous coupled-wave equations. In fact, only the component of the wavevectors along the boundary in the direction of periodicity

(x) needs to satisfy the Floquet condition. This is given by (5) and will be treated below. The Floquet condition for an infinite grating, (4), contains the required x dependence. The z dependence in (4) will be shown to be very useful in obtaining a directly solvable form of the coupled-wave equations. Substituting (4) into (6) yields

$$E(x,z)=\sum_{i=-\infty}^{+\infty} S_i(z)\exp[-j(\bar{k}_2-i\bar{K})\cdot\bar{r}]$$
$$=\sum_{i=-\infty}^{+\infty} S_i(z)\exp\{-j[(k_2\sin\theta-iK\sin\phi)x+(k_2\cos\theta-iK\cos\phi)z]\}, \quad (8)$$

where θ is the angle of refraction of the incident beam from region 1. Thus θ is related to θ' through

$$k_1\sin\theta'=k_2\sin\theta. \quad (9)$$

Substituting (1) and (8) into (3) and performing the indicated differentiations gives

$$\sum_{i=-\infty}^{+\infty}\left\{\frac{\partial^2 S_i(z)}{\partial z^2}-j2(k_2\cos\theta-iK\cos\phi)\frac{\partial S_i(z)}{\partial z}-[(k_2\sin\theta-iK\sin\phi)^2+(k_2\cos\theta-iK\cos\phi)^2]S_i(z)+k^2\varepsilon_0 S_i(z)+\frac{k^2\varepsilon_1}{2}S_{i-1}(z)+\frac{k^2\varepsilon_1}{2}S_{i+1}(z)\right\}\cdot\exp\{-j[(k_2\sin\theta-iK\sin\phi)x+(k_2\cos\theta-iK\cos\phi)z]\}=0. \quad (10)$$

This equation must be satisfied for all values of the variables. Thus the coefficient of each exponential must individually be zero for nontrivial solutions. Using this and the definitions of k, k_2, and K, (10) reduces to the rigorous coupled-wave equations:

$$\frac{1}{2\pi^2}\frac{d^2S_i(z)}{dz^2}-j\frac{2}{\pi}\left[\frac{(\varepsilon_0)^{1/2}\cos\theta}{\lambda}-\frac{i\cos\phi}{\Lambda}\right]\frac{dS_i(z)}{dz}+\frac{2i(m-i)}{\Lambda^2}S_i(z)+\frac{\varepsilon_1}{\lambda^2}[S_{i+1}(z)+S_{i-1}(z)]=0. \quad (11)$$

This is an infinite set of second-order coupled difference-differential equations. By inspection, it is seen that the wave corresponding to each value of i (space harmonic inside the grating or diffracted order outside of the grating) is coupled to its adjacent $(i+1$ and $i-1)$ space harmonics. There is no direct coupling between nonadjacent orders.

In the rigorous coupled-wave equations the quantity m has been defined as

$$m=\frac{2\Lambda(\varepsilon_0)^{1/2}}{\lambda}\cos(\theta-\phi). \quad (12)$$

The quantity m may have any value in general. For the case when m is an integer, (12) becomes the Bragg condition. However, it is important to realize that the Bragg condition is not specifically an input into this theory. The approach applies to an arbitrary angle of incidence and wavelength. Only if the angle of incidence and wavelength are such that m is an integer does Bragg incidence occur. The rigorous coupled-wave equations given by (11) are in a from that is directly solvable using a state variables approach from linear systems theory. This method of solution will be used in the next section.

The Floquet condition as given by (5) may also be used in space harmonic field expansion, (6), to obtain an alternative set of rigorous coupled-wave equations. Substituting (5) into (6) yields

$$E(x,z)=\sum_{i=-\infty}^{+\infty} S_i(z)\exp[-j(k_{2x}-iK_x)x]$$
$$=\sum_{i=-\infty}^{+\infty} S_i(z)\exp[-j(k_2\sin\theta-iK\sin\phi)x]. \quad (13)$$

The field expansion given by (13) is the same as that in (8) except that the z dependent part of the exponential has been included in the $S_i(z)$ functions in (13). Substituting (1) and (13) into (3), performing the indicated differentiations, and setting the coefficients of each exponential equal to zero as before, produces

$$\frac{1}{2\pi^2}\frac{d^2S_i(z)}{dz^2}-2\left\{\left[\frac{(\varepsilon_0)^{1/2}\sin\theta}{\lambda}-\frac{i\sin\phi}{\Lambda}\right]^2-\frac{\varepsilon_0}{\lambda^2}\right\}S_i(z)+\frac{\varepsilon_1}{\lambda^2}\exp\left(+j\frac{2\pi z\cos\phi}{\Lambda}\right)S_{i-1}(z)+\frac{\varepsilon_1}{\lambda^2}\exp\left(-j\frac{2\pi z\cos\phi}{\Lambda}\right)S_{i+1}(z)=0. \quad (14)$$

This set of coupled-wave equations contains no first derivative terms in contrast to the coupled-wave equations given in (11). In addition, (14) is a nonconstant coefficient differential equation due to the presence of z in the coefficients of the $S_{i-1}(z)$ and $S_{i+1}(z)$ terms. The equations in the form of (14) represent a linear shift-variant system and direct solution would be difficult. For the case of an unslanted grating ($\phi=\pi/2$, fringes perpendicular to the surface), the equations become constant coefficient differential equations. For this limiting case, the equations become identical to the coupled-wave equations of Kong [Ref. 6, Eqs. (6a) and (6b)] if only two waves are retained ($i=0, 1$).

4. Solution of the Rigorous Coupled-Wave Equations

The rigorous coupled-wave equations as given by (11) represent a set of second-order linear differential equations with constant coefficients. Using the methods of linear systems analysis [24] this differential equation description of this continuous system may be trans-

formed into a state space description and a solution obtained directly. By defining the state variables as

$$S_{1,i}(z)=S_i(z), \tag{15}$$

$$S_{2,i}(z)=\frac{dS_i(z)}{dz} \tag{16}$$

the infinite set of second-order differential equations (11) are transformed into two infinite sets of first-order differential equations:

$$\frac{dS_{1,i}(z)}{dz}=S_{2,i}(z), \tag{17}$$

$$\frac{dS_{2,i}(z)}{dz}=-\frac{2\pi^2\varepsilon_1}{\lambda^2}S_{1,i-1}(z)+\frac{4\pi^2 i(i-m)}{\Lambda^2}S_{1,i}(z)-\frac{2\pi^2\varepsilon_1}{\lambda^2}S_{1,i+1}(z)+j4\pi\left(\frac{(\varepsilon_0)^{1/2}\cos\theta}{\lambda}-\frac{i\cos\phi}{\Lambda}\right)S_{2,i}(z). \tag{18}$$

Equations (17) and (18) are the state equations corresponding to the rigorous coupled-wave equations (11). Since these are homogeneous equations, they correspond to unforced state equations. State equations that are linear differential equations with constant coefficients such as these, may be solved for closed-form expressions for the state variables. In this case, only the homogeneous solution is necessary as there are no driving terms in these equations. The homogeneous solutions are

$$S_{l,i}(z)=\sum_{m=-\infty}^{+\infty}C_m w_{l,im}\exp(\lambda_m z) \tag{19}$$

for $l=1,2$. The coefficients C_m are unknown constants to be determined from the boundary conditions. The quantity $w_{l,im}$ is an element of an eigenvector and λ_m is an eigenvalue. These quantities are determined as described below. The solution for the wave amplitudes (the "output equation" in linear systems terminology) is $S_i(z)=S_{1,i}(z)$.

The constituent state equations (17) and (18) may be written in matrix form as

$$\begin{bmatrix}\vdots\\ \dot S_{1,-2}\\ \dot S_{1,-1}\\ \dot S_{1,0}\\ \dot S_{1,1}\\ \dot S_{1,2}\\ \vdots\\ \dot S_{2,-2}\\ \dot S_{2,-1}\\ \dot S_{2,0}\\ \dot S_{2,1}\\ \dot S_{2,2}\\ \vdots\end{bmatrix}=\begin{bmatrix}&&&\vdots&&&&&\vdots&&&\\ &0&0&0&0&0&1&0&0&0&0&\\ &0&0&0&0&0&0&1&0&0&0&\\ \cdots&0&0&0&0&0&0&0&1&0&0&\cdots\\ &0&0&0&0&0&0&0&0&1&0&\\ &0&0&0&0&0&0&0&0&0&1&\\ &&&&\vdots&&&&\vdots&&&\\ &b_{-2}&a&0&0&0&c_{-2}&0&0&0&0&\\ &a&b_{-1}&a&0&0&0&c_{-1}&0&0&0&\\ \cdots&0&a&b_0&a&0&0&0&c_0&0&0&\cdots\\ &0&0&a&b_1&a&0&0&0&c_1&0&\\ &0&0&0&a&b_2&0&0&0&0&c_2&\\ &&&&\vdots&&&&\vdots&&&\end{bmatrix}\begin{bmatrix}\vdots\\ S_{1,-2}\\ S_{1,-1}\\ S_{1,0}\\ S_{1,1}\\ S_{1,2}\\ \vdots\\ S_{2,-2}\\ S_{2,-1}\\ S_{2,0}\\ S_{2,1}\\ S_{2,2}\\ \vdots\end{bmatrix}, \tag{20}$$

where $a=-2\pi^2\varepsilon_1/\lambda^2$, $b_i=4\pi^2 i(i-m)/\Lambda^2$, and $c_i=j4\pi[(\varepsilon_0)^{1/2}\cos\theta/\lambda-i\cos\phi/\Lambda]$. This equation may be represented concisely as $\dot{\mathbf{S}}=\mathbf{AS}$ where $\dot{\mathbf{S}}$ and $\mathbf{S}$ are the column vectors in (20) and $\mathbf{A}$ is the coefficient matrix. The needed eigenvalues and eigenvectors are determined from this coefficient matrix. Although $\mathbf{A}$ is an infinite matrix, results may be obtained in practice to an arbitrary level of accuracy with a truncated matrix. Each of the four submatrices is truncated to $n\times n$. As the integer n increases, the calculated results rapidly converge to the exact results. The quantity n corresponds to the total number of space harmonics retained in the analysis. This in turn means that the analysis includes n diffracted waves in region 1 and n diffracted waves in region 3. To put the four submatrices into standard form, the integers i and m are replaced with the new integers p and q that run from 1 to n. For example, if an odd number of waves are retained symmetrically about $i=0$ (the undiffracted transmitted wave) in the analysis, then $p=i+(n+1)/2$ and $q=m+(n+1)/2$. The $2n$ solutions may then be expressed

$$S_{l,p}(z)=\sum_{r=1}^{2}\sum_{q=1}^{n}C_{r,q}w_{l,p;r,q}\exp(\lambda_{r,q}z) \tag{21}$$

for $l=1,2$ and $p=1$ to n. The eigenvalues $\lambda_{r,q}$ are determined by solving the determinantal equation

$$|\mathbf{A}-\lambda_{r,q}\mathbf{I}|=0, \tag{22}$$

where $\mathbf{I}$ is the unit matrix. The eigenvector corresponding to a particular eigenvalue $\lambda_{r,q}$ is determined by substituting $2n$ expressions ($l=1,2$ and $p=1$ to n) for $S_{l,p}$ of the form $S_{l,p}=B_{l,p;r,q}\exp(\lambda_{r,q})$ into the state equation (20), performing the indicated differentiations, and then solving for each element of the eigenvector as $w_{l,p;r,q}=B_{l,p;r,q}/B_{1,1\pi;r,q}$ using Cramer's rule and thus expressing each element as a ratio of determinants. The eigenvalues and eigenvectors for a

matrix are typically calculated numerically using a computer library program [25].

5. Equivalence of Coupled-Wave and Modal Representations

The total field inside the grating may be expressed in coupled-wave form, (6), or in modal form, (7), depending whether the field is expanded in terms of space harmonics or in terms of modes, respectively. These two forms are alternative representations that are completely equivalent. This equivalence has been discussed previously [14, 19, 21]. It can be shown mathematically in a simple manner as follows.

Substituting $S_i(z)=\sum_{m=-\infty}^{+\infty} C_m w_{1,im}\exp(\lambda_m z)$ into the coupled-wave expansion (8) gives

$$E(x,z)=\sum_{i=-\infty}^{+\infty}\sum_{m=-\infty}^{+\infty} C_m w_{1,im} \cdot\exp\{-j[(k^2\sin\theta-iK\sin\phi)x +(k_2\cos\theta-iK\cos\phi+j\lambda_m)z]\}. \tag{23}$$

Changing the order of the summation this may be rewritten as

$$E(x,z)=\sum_{m=-\infty}^{+\infty} C_m \sum_{i=-\infty}^{+\infty} w_{1,im}\exp[-j(\bar{k}_{2m}-i\bar{K})\cdot\bar{r}], \tag{24}$$

where $\bar{k}_{2m}=k_2\sin\theta\,\hat{x}+(k_2\cos\theta+j\lambda_m)\hat{z}$. Identifying the complex Fourier series and its representation of a periodic function

$$\sum_{i=-\infty}^{+\infty} w_{1,im}\exp(ji\bar{K}\cdot\bar{r})=\Phi_m(\bar{r})=\Phi_m(\bar{r}+\bar{\Lambda}) \tag{25}$$

gives the modal expansion

$$E(x,z)=\sum_{m=-\infty}^{+\infty} C_m\Phi_m(\bar{r})\exp(-j\bar{k}_{2m}\cdot\bar{r}) \tag{26}$$

which is identical to (7). Thus the coupled-wave and modal representations are seen to be equivalent.

6. Phase Matching and Boundary Conditions

Each i-th field in region 1 and 3 must be phase matched to the i-th space harmonic field inside the grating. In addition, the magnitude of the fields in regions 1 and 3 must be such that the electromagnetic boundary conditions are satisfied at the two grating boundaries ($z=0$ and $z=d$).

The total electric field in region 1 is the sum of the incident and the backward-traveling waves. The normalized total electric field in region 1 may be expressed as

$$E_1=\exp(-j\bar{k}_1\cdot\bar{r})+\sum_{i=-\infty}^{\infty} R_i\exp(-j\bar{k}_{1i}\cdot\bar{r}), \tag{27}$$

where R_i is the normalized amplitude of the i-th reflected wave in region 1 with wavevector $\bar{k}_{1i}$. Likewise the normalized total electric field in region 3 is

$$E_3=\sum_{i=-\infty}^{\infty} T_i\exp[-j\bar{k}_{3i}\cdot(\bar{r}-d\hat{z})], \tag{28}$$

where T_i is the normalized amplitude of the i-th transmitted wave in region 3 with wavevector $\bar{k}_{3i}$. These fields in regions 1 and 3 are phased matched to the field in the grating, (8). Thus the x components of the wavevectors of the i-th wave (regions 1 and 3) and the x component of the wavevector of the i-th space harmonic field (region 2) must be the same. That is

$$\bar{k}_{1i}\cdot\hat{x}=(\bar{k}_2-i\bar{K})\cdot\hat{x}=\bar{k}_{3i}\cdot\hat{x}$$
$$k_1\sin\theta_i'=k_2\sin\theta-iK\sin\phi=k_3\sin\theta_i''. \tag{29}$$

In the homogeneous regions (1 and 3) the backward- and forward-diffracted waves have wavevectors with magnitudes

$$|\bar{k}_{1i}|=|\bar{k}_1| \quad\text{and}\quad |\bar{k}_{3i}|=|\bar{k}_3|, \tag{30}$$

where $k_3=2\pi(\varepsilon_{III})^{1/2}/\lambda$. Knowing the total amplitudes and the x components of the diffracted wavevectors, the z components are then determined to be

$$\bar{k}_{1i}\cdot\hat{z}=[|\bar{k}_1|^2-(\bar{k}_{1i}\cdot\hat{x})^2]^{1/2} =[k_1^2-(k_2\sin\theta-iK\sin\phi)^2]^{1/2} \tag{31}$$

and

$$\bar{k}_{3i}\cdot\hat{z}=[|\bar{k}_3|^2-(\bar{k}_{3i}\cdot\hat{x})^2]^{1/2} =[k_3^2-(k_2\sin\theta-iK\sin\phi)^2]^{1/2}. \tag{32}$$

These quantities are either positive real (propagating wave) or negative imaginary (evanescent wave).

Using the phase matching conditions (29), (31), and (32), the total fields in regions 1 and 3, (27) and (28), may be rewritten

$$E_1=\exp\{-j[k_1(\sin\theta' x+\cos\theta' z)]\} +\sum_{i=-\infty}^{\infty} R_i\exp[-j\{(k_2\sin\theta-iK\sin\phi)x -[k_1^2-(k_2\sin\theta-iK\sin\phi)^2]^{1/2}z\}] \tag{33}$$

and

$$E_3=\sum_{i=-\infty}^{\infty} T_i\exp\{-j\{(k_2\sin\theta-iK\sin\phi)x +[k_3^2-(k_2\sin\theta-iK\sin\phi)^2]^{1/2}(z-d)\}\}. \tag{34}$$

Electromagnetic boundary conditions require that the tangential electric and tangential magnetic fields be continuous across the two boundaries ($z=0$ and $z=d$). For the H mode polarization described in this paper, the electric field only has a component in the y

direction and so it is the tangential electric field directly. The magnetic field intensity, however, must be obtained through the Maxwell equation $\nabla x\bar{E} = -\partial\bar{B}/\partial t$. The tangential component of H is in the x direction and is thus given by $H_x = (-j/\omega\mu)\partial E_y/\partial z$. For each value of i, the four quantities to be matched and the resulting boundary condition are:

1) tangential E at $z=0$:

$$\delta_{i0} + R_i = S_i(0), \tag{35}$$

2) tangential H at $z=0$:

$$j[k_1^2 - (k_2 \sin\theta - iK\sin\phi)^2]^{1/2}(R_i - \delta_{i0}) = \frac{dS_i(0)}{dz}, \tag{36}$$

3) tangential E at $z=d$:

$$T_i = S_i(d)\exp[-j(k_2\cos\theta - iK\cos\phi)d], \tag{37}$$

4) tangential H at $z=d$:

$$-j[k_3^2 - (k_2\sin\theta - iK\sin\phi)^2]^{1/2} T_i = \left[\frac{dS_i(d)}{dz} - j(k_2\cos\theta - iK\cos\phi)S_i(d)\right] \cdot \exp[-j(k_2\cos\theta - iK\cos\phi)d], \tag{38}$$

where δ_{i0} is the Kronecker delta function.

7. Diffraction Efficiency

The quantity commonly measured in grating diffraction is the diffraction efficiency. It is defined as the diffracted intensity of a particular order divided by the input intensity. In the above formulation, the incident plane wave amplitude was normalized to unity. Thus the diffraction efficiencies in regions 1 and 3 are

$$\begin{aligned} \mathrm{DE}_{1i} &= \mathrm{Re}\{(\bar{k}_{1i}\cdot\hat{z})/(\bar{k}_{10}\cdot\hat{z})\} R_i R_i^* \\ &= \mathrm{Re}\{\{1 - [\sin\theta' - i\lambda\sin\phi/(\varepsilon_1)^{1/2}\Lambda]^2\}^{1/2}/ \\ &\quad \cos\theta'\} R_i R_i^* \end{aligned} \tag{39}$$

and

$$\begin{aligned} \mathrm{DE}_{3i} &= \mathrm{Re}\{(\bar{k}_{3i}\cdot\hat{z})/(\bar{k}_{10}\cdot\hat{z})\} T_i T_i^* \\ &= \mathrm{Re}\{\{(\varepsilon_{\mathrm{III}}/\varepsilon_1) - [\sin\theta' - i\lambda\sin\phi/(\varepsilon_1)^{1/2}\Lambda]^2\}^{1/2}/ \\ &\quad \cos\theta'\} T_i T_i^*. \end{aligned} \tag{40}$$

The real part of the ratio of the propagation constants occurs when the time-average power-flow density is obtained by taking the real part of the complex Poynting vector. For an unslanted grating ($\phi = \pi/2$) with the same medium on both sides ($\varepsilon_{\mathrm{I}} = \varepsilon_{\mathrm{III}}$), the real part of the ratio of the propagation constants is just the usual ratio of the cosine of the diffraction angle for the i-th diffracted wave to the cosine of the incidence angle.

If n values of i are retained in the analysis, then there will be n forward-diffracted waves (n values of T_i) and n backward-diffracted waves (n values of R_i). Correspondingly, there will be $2n$ unknown values of C_{m}. This is because the coefficient matrix in (20) is a $2n \times 2n$ matrix and therefore has $2n$ eigenvalues and thus there are $2n$ unknown values of C_{m}. Also this may be viewed as being due to the n coupled-wave equations, each being a second-order differential equation, and thus there are $2n$ roots or eigenvalues and $2n$ unknown constants C_{m} to be determined from the boundary conditions. Therefore, the total number of unknowns is $4n$. Substituting $S_i(z)$, as given by (15) and (19), into the equations for the boundary conditions (35)–(38) produces n linear equations containing the $4n$ unknowns. An efficient procedure to solve these equations is to eliminate R_i and T_i from these equations and to solve the resulting $2n$ equations for the $2n$ values of C_{m} using a technique such as Gauss elimination. Then the n values of R_i and n values of T_i may be determined from (35) and (37) respectively. Finally, the diffraction efficiencies DE_{1i} and DE_{3i} are calculated using (39) and (40). For phase gratings the input power is conserved and thus the sum of all of the efficiencies for the propagating waves is unity. That is,

$$\sum_i (\mathrm{DE}_{1i} + \mathrm{DE}_{3i}) = 1. \tag{41}$$

Equation (41) may be used to verify the convergence of the numerical calculations.

8. Approximate Theories

The vast majority of the papers on grating diffraction theory have dealt with approximate theories. There are a large number of possible approximations and assumptions that can be made. These generally lead to enormous simplifications in the analyses. In some cases, these simplifications allow analytic solutions to be obtained. A number of famous analytic expressions occur for special limiting cases.

In this section, a large number of planar grating diffraction theories are classified in terms of the fundamental assumptions: 1) neglect of higher-order waves, 2) neglect of second derivatives of the field amplitudes, 3) neglect of boundary effects, 4) neglect of dephasing from the Bragg condition, and 5) the small grating modulation approximation. In addition to these assumptions, a number of other approximations such as normal incidence, unslanted gratings, and large grating period compared to a wavelength, may also be made. However, in this section, only the fundamental assumptions enumerated above are treated. Thus all of the approximate theories are presented in their general form allowing for arbitrary angle of incidence (θ'),

arbitrary grating slant angle (ϕ), and arbitrary grating period (Λ). The various further reductions can then be easily formulated, if desired, from these general forms of the approximate theories.

In region 1 of Fig. 1, backward-traveling waves exist. In general, these waves are produced both by diffraction from within the grating volume and by boundary effects (diffraction and reflection from the periodic boundaries at $z=0$ and $z=d$). These physical processes produce a spectrum of plane waves traveling back into region 1 ($z<0$). For the general planar grating of Fig. 1, neglecting the second derivatives of the field amplitudes in the wave equation reduces the number of waves in the analysis from $2n$ to n. The bulk diffracted orders are retained and the boundary-produced waves are eliminated. Thus for a planar grating, the neglect of second derivatives and the neglect of boundary effects are absolutely linked together. When these assumptions are made, the resulting first-order coupled-wave analyses have the amplitudes of the diffracted waves calculated *inside* the modulated region. Then the amplitudes T_i of the forward-diffracted output waves are obtained (approximately) by arguing that they are equal to $S_i(d)$, the space harmonic field amplitude at a distance d from the input surface $z=0$. Likewise for those values of i that represent backward-diffracted waves, the amplitudes R_i are estimated to be $S_i(0)$. However, in the physical problem being analyzed, there are no boundaries at $z=0$ and $z=d$. These planes just represent reference locations. There are no reflected or diffracted waves resulting from these planes and thus there are no physical boundaries at these locations! Thus, the assumptions of neglecting the second derivatives of field amplitudes and neglecting boundary effects have transformed the problem into a filled-space problem (a grating filling all space) with imaginary boundaries at $z=0$ and $z=d$ that are used only to obtain an approximate mathematical formulation of the problem. The first-order theory approaches are not capable of solving the problem of the general planar slab grating bounded by two media different from the grating medium. These two linked assumptions therefore, unmistakably imply the filled-space problem. *After* the filled-space problem is solved, *then* it is assumed that the grating terminates at $z=0$ and $z=d$ and, as a result, that $T_i \cong S_i(d)$ for the forward-diffracted waves and $R_i \cong S_i(0)$ for the backward-diffracted waves. This is obviously only an approximation to the actual situation.

Another consequence of neglecting second derivatives is the exclusion of some propagating waves. In first-order theory, only half of the waves can be retained in the analysis. That is, only one set of i values (as opposed to two sets) is included. For a genral slanted grating, some of these waves may be forward-diffracted and some of them may be backward-diffracted. From (8), if $k_2 \cos\theta - iK\cos\phi$ is positive, the wave is forward-diffracted and if negative, it is backward-diffracted. For forward-diffracted waves, the boundary condition used must be $S_i(0)=0$. For backward-diffracted waves, the appropriate boundary condition is $S_i(d)=0$. The second set of waves (set of i values) are phase matched to these waves. This second set of waves is, of course, neglected in any first order analysis. For the example depicted in Fig. 2, the backward-diffracted waves for $-1 \leqq i \leqq +4$ would all be neglected in first-order theory. The diffraction efficiencies of these backward-diffracted waves are arbitrarily set equal to zero. For the case of a slanted-fringe grating, the power in the neglected phase-matched waves has been shown to be very significant in some cases [8]. Thus the errors introduced by using first order theory can be particularly significant for slant angles away from $\phi=0$ and $\phi=\pi/2$.

Still another consequence of neglecting second derivatives is the exclusion of evanescent waves from the analysis. In first order theory, the filled-space nature of the grating being analyzed, causes one complete set of diffracted orders (i) to exist inside the grating, since all of the $S_i(z)$'s exist there. These calculated values of $S_i(z)$ may have wavevectors with components either in the $+$ or $-$ z directions. However, many of the wavevectors of the $S_i(z)$'s cannot be phased matched to plane waves outside of the grating (regions 1 and 3). This may be seen from Fig. 2. For this example, the values $-1 \leqq i \leqq +4$ correspond to propagating plane waves in regions 1 and 3. The values $i \leqq -2$ and $i \geqq +5$ correspond to evanescent waves in region 1 and 3. However, in first order analysis (without second derivatives) all values of i are treated as representing propagating waves. This is obviously not true. Nevertheless, diffraction efficiencies can be calculated for these evanescent waves as though they were propagating. These predicted efficiencies are clearly incorrect since they should be zero. If the grating period is much larger than a wavelength ($\Lambda \gg \lambda$), then there will be a large number of propagating waves and the effect of excluding evanescent waves would be reduced.

Therefore, it is concluded that all first-order theories inherently contain: 1) the approximate method for calculating diffracted amplitudes described above, 2) neglect of phase-matched waves, and 3) neglect of evanescent waves.

A depiction of various planar grating diffraction theories and their interrelationships in terms of fundamental assumptions is shown in Fig. 3. Most of the literature on planar grating diffraction theory can be connected with a particular block in this diagram. The importance of the various assumptions cannot always

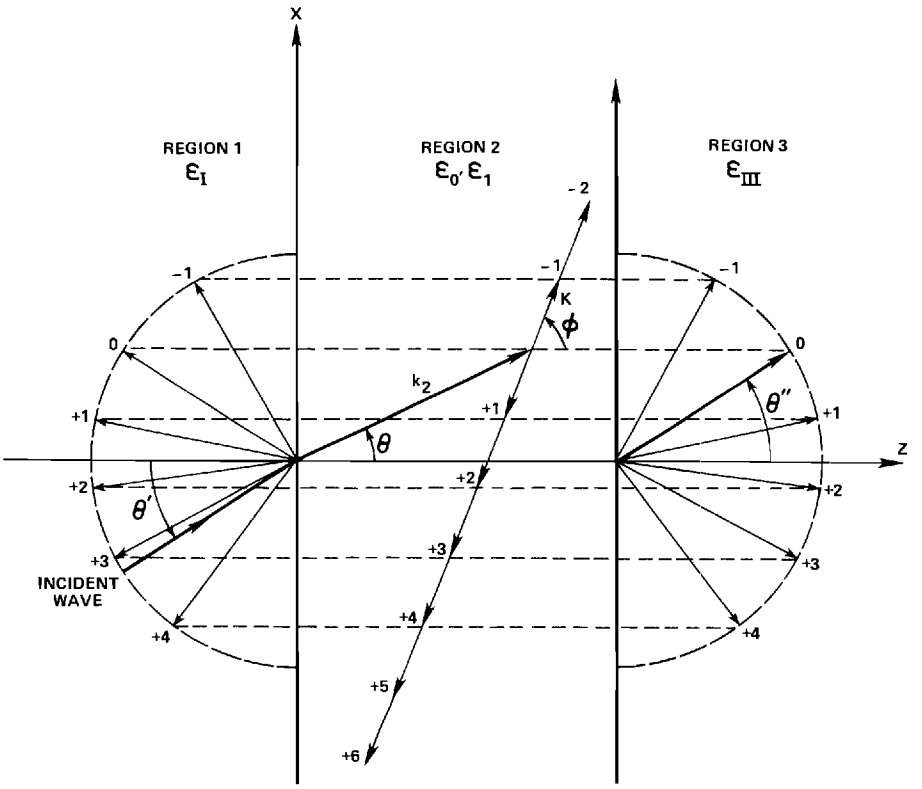


Fig. 2. Allowed wavevectors in regions 1 and 3 due to the presence of a slanted grating with wavevector $\bar{K}$. Phase matching of the diffracted waves outside of the grating with the boundary components of the wavevectors inside the grating (via Floquet construction) is shown. For $-1 \leqq i \leqq +4$, propagating diffracted orders exist in regions 1 and 3 whereas for $i \leqq -2$ and $i \geqq +5$, the waves are evanescent

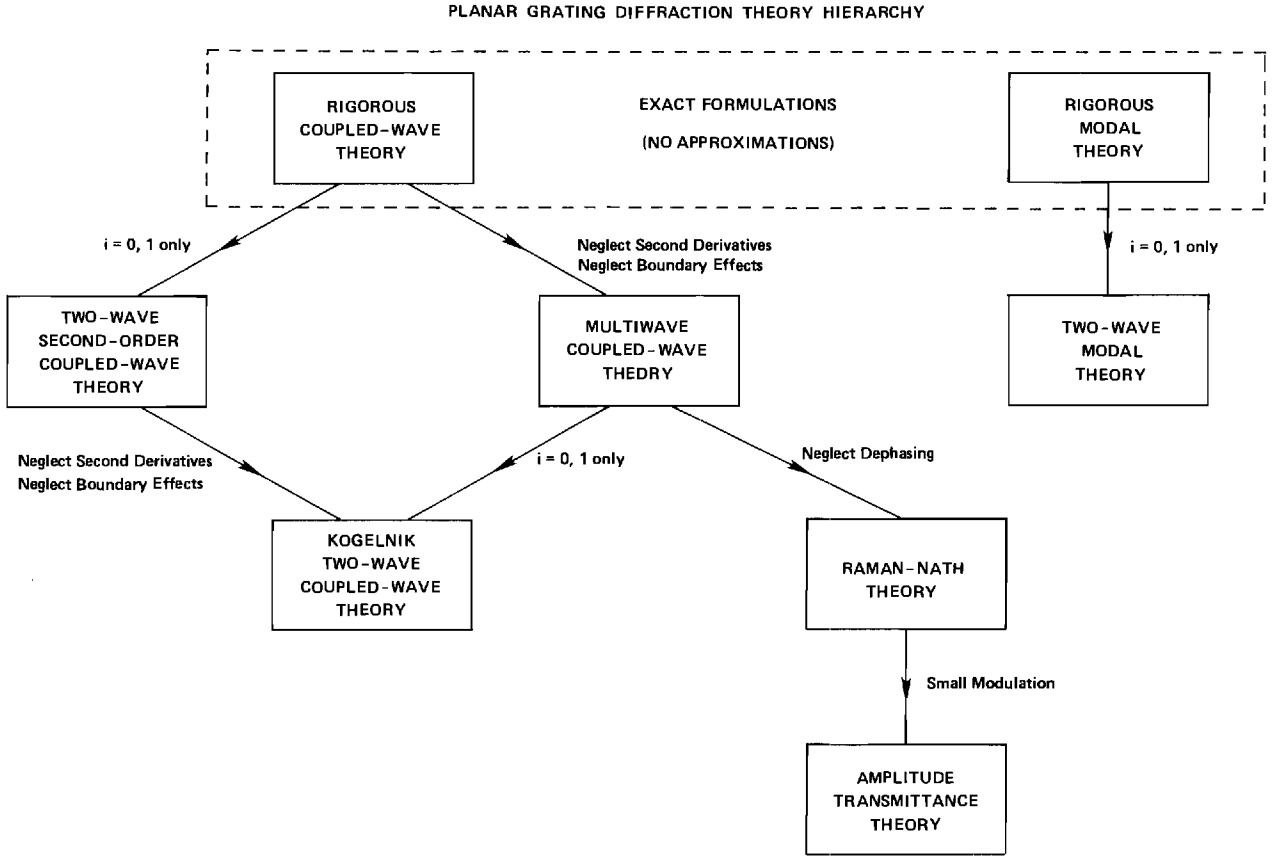


Fig. 3. Interrelationships between various planar grating diffraction theories in terms of fundamental approximations

be isolated. For example, retaining only two waves ($i=0, 1$) and the small modulation approximation can be linked for the case of a sufficiently "thick" grating.

8.1. Two-Wave Modal Theory

If only the zero and first order waves ($i=0, 1$) are retained and all higher-order waves are neglected, a two-wave regime is being assumed. There are actually a total of four waves in this analysis since there are two more waves phased matched to these. Modal theory solutions in the two-wave regime were first obtained by Bergstein and Kermisch [13] with more recent results being contributed by Lederer and Langbein [26], and Russell [19]. In this approch, the standard modal expansion (7) is used to represent the fields in the grating. However, in the two-wave case only the first two Fourier components ($i=0, 1$) of the periodic function $\Phi_m(\bar{r})$ are retained in the analysis, (25). Comparison of two-wave modal theory with exact rigorous theory [8] has shown that this can be valid near Bragg incidence in reflection gratings (backward-diffracted waves dominate). Comparison data are shown in [Ref. 8, Fig. 9].

8.2. Two-Wave Second-Order Coupled-Wave Theory

Two-wave second-order coupled-wave theory and two-wave modal theory represent exactly the same approximation. Both representations include second derivatives of field amplitudes and boundary effects. Both theories retain only the transmitted wave ($i=0$) and the fundamental diffracted wave ($i=1$) and their phased matched waves and neglect higher-order waves. This approximate theory has been used by Kong [6]. Additional approximations in this theory have been made by Kessler and Kowarschik [27–29], and by Jaaskelainen et al. [30]. The two governing equations may be obtained directly from the rigorous coupled-wave equations (11) by keeping only terms in S_0 and S_1 and neglecting all other field amplitudes. The resulting two equations from (11) are:

$$\frac{1}{2\pi^2}\frac{d^2S_0(z)}{dz^2} - j\frac{2(\varepsilon_0)^{1/2}\cos\theta}{\pi\lambda}\frac{dS_0(z)}{dz} + \frac{\varepsilon_1}{\lambda^2}S_1(z) = 0, \quad (42)$$

$$\frac{1}{2\pi^2}\frac{d^2S_1(z)}{dz^2} - j\frac{2}{\pi}\left[\frac{(\varepsilon_0)^{1/2}\cos\theta}{\lambda} - \frac{\cos\phi}{\Lambda}\right]\frac{dS_1(z)}{dz} + \frac{2(m-1)}{\Lambda^2}S_1(z) + \frac{\varepsilon_1}{\lambda^2}S_0(z) = 0. \quad (43)$$

Kong [6] has presented analytical solutions for the two-wave second-order coupled-wave theory expressed in the form of two transmission and two reflection coefficients for the unslanted-fringe planar slab grating.

8.3. Multiwave Coupled-Wave Theory

Multiwave first-order coupled-wave theory may also be developed directly from the rigorous coupled-wave equations (11). In this approach higher-order waves are retained (hence "multiwave"). The second derivatives of the field amplitudes (and thus boundary effects) are neglected. The resulting multiwave coupled-wave equations from (11) are:

$$-j\frac{2}{\pi}\left[\frac{(\varepsilon_0)^{1/2}\cos\theta}{\lambda} - \frac{i\cos\phi}{\Lambda}\right]\frac{dS_i(z)}{dz} + \frac{2i(m-i)}{\Lambda^2}S_i(z) + \frac{\varepsilon_1}{\lambda^2}[S_{i+1}(z)+S_{i-1}(z)] = 0. \quad (44)$$

For the case of an unslanted transmission grating ($\phi=\pi/2$) and normal incidence ($\theta=0$, $m=0$), the multiwave coupled-wave equations first appeared in a 1936 paper by Raman and Nath [31] for a sinusoidal (rather than cosinusoidal) grating. This paper was the fourth in a series of five papers by Raman and Nath [31–35] on the diffraction of light by sound waves. The first three papers [32–34] from the basis of the "Raman-Nath theory" described below. This simplified multiwave coupled-wave equation was referred to by Nath [36] as being due to Nath [37]. In this 1936 paper, Nath [37] obtained a very slowly converging series solution for the multi-wave coupled-wave difference-differential equations. An alternative series solution was later presented by Berry [38]. This series solution is in terms of Bessel functions and is also very slowly converging. Numerical solutions of the multiwave coupled-wave equations (also for acousto-optic interaction studies) have been obtained by Klein and Cook [3].

The multiwave coupled-wave equations have been generalized to include loss and gratings of arbitrary nonsinusoidal profile by Magnusson and Gaylord [7]. In that paper, numerical solutions were obtained for unslanted transmission gratings using a Runge-Kutta algorithm to solve the first-order system of coupled-wave equations. Diffraction efficiency results for sinusoidal, square-wave, and sawtooth phase gratings at first, second, and third Bragg incidence are presented there.

Comparison of diffraction efficiency results from multiwave first-order coupled-wave theory with exact rigorous theory has shown that this theory without second derivatives gives good results in transmission gratings (forward-diffraction waves dominate) when the grating modulation is small. Comparison data are shown in [Ref. 8, Figs. 7 and 8].

8.4. Two-Wave First-Order Coupled-Wave Theory

If higher-order waves ($i\neq 0, 1$) and second derivatives of field amplitudes (and thus boundary effects) are *both*

neglected, the rigorous coupled-wave equations(11) reduce to two-wave first-order coupled-wave theory. For general slanted gratings at arbitrary incidence the two governing equations are:

$$\cos\theta\frac{dS_0(z)}{dz}+j\frac{\pi\varepsilon_1}{2(\varepsilon_0)^{1/2}\lambda}S_1(z)=0, \tag{45}$$

$$\left(\cos\theta-\frac{\lambda}{\Lambda(\varepsilon_0)^{1/2}}\right)\frac{dS_1(z)}{dz}+j\frac{\pi\lambda(m-1)}{\Lambda^2(\varepsilon_0)^{1/2}}S_1(z)$$
$$+j\frac{\pi\varepsilon_1}{2(\varepsilon_0)^{1/2}\lambda}S_0(z)=0. \tag{46}$$

Two-wave first-order coupled-wave theory was applied to acousto-optics by Phariseau [2]. It was first applied to holography by Kogelnik [4]. His thorough 1969 paper [4] is now very widely referenced. As a result, this theory is commonly called "Kogelnik theory" and this is noted in Fig. 3. The substantial recognition received by Kogelnik's paper [4] is due in part to the comprehensive coverage of 1) phase, absorption, and mixed gratings; 2) on-Bragg and off-Bragg incidence; 3) pure transmission ($\phi=\pi/2$), pure reflection ($\phi=0$), and general slanted fringe gratings; and 4) both H-mode and E-mode polarization.

From (8), if $k_2\cos\theta-K\cos\phi$ is positive, the single diffracted wave in this analysis is forward-diffracted and the grating is called a transmission grating. If $k_2\cos\theta-K\cos\phi$ is negative, the single diffracted wave is backward-diffracted and the grating is called a reflection grating. For the forward-diffracted case, the boundary condition used is $S_1(0)=0$. In the backward-diffracted case, the boundary condition used is $S_1(d)=0$. Due to the first-order nature of this theory, some phase matched waves will be neglected. In the transmission grating case, for example, the two backward-traveling waves (that are phase matched to the zero-order transmitted wave and the fundamental diffracted wave) are neglected.

For the special case of a phase grating with unslanted fringes ($\phi=\pi/2$) and incidence at the first Bragg angle ($m=1$), the first-order diffracted amplitude from (45) and (46) is given by [2, 4]

$$S_1(z)=-j\sin\left(\frac{\pi\varepsilon_1 z}{2(\varepsilon_0)^{1/2}\lambda\cos\theta}\right), \tag{47}$$

where z is the distance into the grating at which the amplitude is determined. This well-known expression predicts a diffraction efficiency [$\mathrm{DE}=S_i(d)S_i^*(d)$ for this case] that is sinusoidal in modulation and has a maximum value of 100%. Although the two-wave first-order coupled-wave theory neglects higher-order diffracted waves and second derivatives of field amplitudes (and thus also boundary effects), it nevertheless contains many of the basic features of the diffraction process in an extended grating. This theory has been sucessfully extended to numerous other cases including finite beams [39, 40], finite and nonplanar gratings [41–43], and attenuated gratings [29, 44–46]. When grating diffraction is described by the two-wave result, (47), it is often referred to as "Bragg regime" diffraction. Incidence at the Bragg angle is essential in "Bragg regime" diffraction whereas in "Raman Nath regime" diffraction described below it is not. Criteria for "Bragg regime" behavior are given in [47].

For the basic case of a uniform grating and plane wave, a comparison of diffraction efficiency results from two-wave first-order coupled-wave theory with exact rigorous theory is presented in [8] for a series of grating slant angles. When transmission grating behavior dominates, the error in two-wave coupled-wave theory. is due primarily to the neglect of higher-order waves in the theory. Conversely, when reflection grating behavior dominates, the error is primarily due to the neglect of second derivatives and boundary effects.

8.5. Raman-Nath Theory

The theory of Raman and Nath [32–35] may also be obtained directly from the rigorous coupled-wave equations. If second derivatives of the field amplitudes and dephasing from the Bragg condition are both neglected, the rigorous coupled-wave equations (11) reduce to the Raman-Nath diffraction equations:

$$-j\frac{2}{\pi}\left[\frac{(\varepsilon_0)^{1/2}\cos\theta}{\lambda}-\frac{i\cos\phi}{\Lambda}\right]\frac{dS_i(z)}{dz}$$
$$+\frac{\varepsilon_1}{\lambda^2}[S_{i+1}(z)+S_{i-1}(z)]=0, \tag{48}$$

where a general angle of incidence and grating slant angle have been retained. The S_i term in (11) has been neglected. For the i-th diffracted order, this term is zero for the m-th Bragg incidence, (12), when $i=m$. For an arbitrary angle of incidence, each diffracted order will be dephased by varying amounts from their corresponding Bragg conditions. This in turn reduces the synchronism between the input wave and that diffracted order. The result is less coupling from the input to that order. The Raman-Nath theory therefore, treats all diffracted orders as though the Bragg conditions for all them were simultaneously satisfield.

For the important case of an unslanted fringe transmission grating ($\phi=\pi/2$), (48) takes the form of a recurrence relation satisfied by Bessel functions. The solution is

$$S_i(z)=(-j)^i\mathrm{J}_i\left(\frac{\pi\varepsilon_1 z}{(\varepsilon_0)^{1/2}\lambda\cos\theta}\right) \tag{49}$$

for boundary conditions $S_0(0)=1$ and $S_i(0)=0$ ($i\neq 0$) where J_i is an integer-order Bessel function of the

first kind. Equation (49) is the famous Bessel function expression of Raman and Nath. It predicts maximum diffraction efficiencies of $DE_{\pm 1}=33.8\%$, $DE_{\pm 2}=23.6\%$, $DE_{\pm 3}=18.8\%$, and so forth. When grating diffraction behavior may be approximated by (49), it is referred to as "Raman-Nath regime" diffraction. This result, (49), has been extensively used to predict the light intensities diffracted by sound waves [38, 48]. Criteria for "Raman-Nath regime" diffraction are given in [49]. Raman-Nath theory has been extended to describe nonsinusoidal phase gratings [50–52].

8.6. *Amplitude Transmittance Theory*

For gratings, the amplitude transmittance approach is closely related to Raman-Nath diffraction theory. The amplitude transmittance approach is widely used in optics [53–54] and may be applied to slabs, lenses, apertures, and general two-dimensional objects as well as gratings. The amplitude transmittance is defined as the ratio of the field amplitude over the output plane to the field amplitude incident on the input plane. The amplitude transmittance function in general is complex. It may be applied to gratings with unslanted fringes. Both amplitude gratings [53–57] and phase gratings [51–56] have been treated in the literature using the amplitude transmittance approach.

For a phase grating with periodicity in the x direction, the amplitude transmittance function is

$$\tau(x,z)=\exp\left(-j\frac{2\pi[\varepsilon(x)]^{1/2}z}{\lambda\cos\theta}\right), \tag{50}$$

where z is the grating thickness and $n(x)=[\varepsilon(x)]^{1/2}$ is the periodic refractive index. Since the transmittance function is also periodic in x, it may be expanded in a complex Fourier series. Further, because the exponentials in this series are in the form of an expansion of the diffracted plane waves, then the Fourier coefficients are the diffracted wave amplitudes. The Fourier series expansion may thus be written

$$\tau(x,z)=\sum_i S_i(z)\exp(\mathrm{j}iKx), \tag{51}$$

where S_i represents the amplitude of the i-th diffracted order. By definition, the coefficients of the Fourier series may be calculated from

$$S_i(z)=\frac{1}{\Lambda}\int_0^{\Lambda}\exp\left(-\mathrm{j}\frac{2\pi[\varepsilon(x)]^{1/2}z}{\lambda\cos\theta}\right)\exp(-\mathrm{j}iKx)dx. \tag{52}$$

Thus the diffracted amplitudes may be determined directly knowing $\varepsilon(x)$ by integrating (52). Results for sinusoidal, square-wave, sawtooth, triangular, and rectangular refractive-index profiles are given in [51].

For the unslanted-fringe (co)sinusoidal-permittivity transmission grating, the corresponding index of refraction is

$$n(x)\equiv[\varepsilon(x)]^{1/2}=(\varepsilon_0+\varepsilon_1\cos Kx)^{1/2} \tag{53}$$

which may be expanded in a Fourier cosine series as

$$[\varepsilon(x)]^{1/2}=[\varepsilon(x)]_0^{1/2}+\sum_{h=1}^{\infty}[\varepsilon(x)]_h^{1/2}\cos(hKx) \tag{54}$$

with Fourier harmonic amplitudes given by

$$[\varepsilon(x)]_h^{1/2}=\frac{2}{\Lambda}\int_0^{\Lambda}(\varepsilon_0+\varepsilon_1\cos Kx)^{1/2}\cos(hKx)dx. \tag{55}$$

The average value of the refractive index may be expressed concisely as

$$n_0(x)\equiv[\varepsilon(x)]_0^{1/2}=(2/\pi)(\varepsilon_0+\varepsilon_1)^{1/2}E(\zeta,\pi/2), \tag{56}$$

where $E(\zeta,\pi/2)$ is the complete elliptic integral of the second kind and $\zeta\equiv 2\varepsilon_1/(\varepsilon_0+\varepsilon_1)$. Clearly, the case of a sinusoidal permittivity (or dielectric constant) being treated throughout this paper is not the same as a sinusoidal refractive-index grating. The index of refraction corresponding to sinusoidal permittivity has higher spatial frequency harmonics ($h>1$) in addition to a fundamental sinusoidal component ($h=1$) as represented by (54). However, for the case of sufficiently small modulation, a sinusoidal permittivity produces nearly a sinusoidal index of refraction. In the limit of small modulation (ε_1 approaches zero), (55) and (56) yield

$$[\varepsilon(x)]_0^{1/2}\simeq\varepsilon_0, \tag{57}$$

$$[\varepsilon(x)]_1^{1/2}\simeq\varepsilon_1/2(\varepsilon_0)^{1/2}, \tag{58}$$

$$[\varepsilon(x)]_2^{1/2},\quad [\varepsilon(x)]_3^{1/2},\ldots\simeq 0. \tag{59}$$

This analysis is important in that it now allows the Raman-Nath theory and amplitude transmittance theory to be interrelated. The result is that although (52) was obtained using the amplitude transmittance approach, it is *also* a solution of the Raman-Nath difference-differential equation (48) for unslanted gratings in the limit of small modulation. This may be shown by direct substitution of S_i as given by (52) into the Raman-Nath diffraction equation (48). Thus, for a cosinusoidal refractive-index profile, the integral in (52) when evaluated gives the Bessel function result (49). This may be accomplished using the identity

$$\exp(-\mathrm{j}b\cos\alpha)\equiv\sum_{i=-\infty}^{+\infty}(-\mathrm{j})^iJ_i(b)\exp(\mathrm{j}i\alpha) \tag{60}$$

and the orthogonality relationship

$$\frac{1}{\Lambda}\int_0^{\Lambda}\exp(\mathrm{j}lKx)\exp(-\mathrm{j}iKx)dx=\delta_{li}, \tag{61}$$

where δ_{li} is the Kronecker delta. Therefore, as depicted in Fig. 3, it has been shown that Raman-Nath theory and amplitude transmittance theory are equivalent in the limit of small grating modulation. This is true in general for unslanted gratings regardless of the grating profile (squarewave, sawtooth, etc.).

9. Discussion and Conclusions

Theories describing the diffraction of a plane electromagnetic wave incident with arbitrary wavelength and angle of incidence upon a slanted fringe planar sinusoidal permittivity grating have been reviewed. For simplicity the analysis has been restricted to H-mode polarization (electric field perpendicular to plane of incidence). Exact formulations without approximations (rigorous coupled-wave analysis and rigorous modal analysis) have been developed and shown to be mathematically equivalent. Rigorous coupled-wave equations have been developed in alternative forms, (11) and (14), and their usefulness discussed. The solution of the rigorous coupled-wave equations (11) in terms of state variables has been presented in detail along with the phase matching and boundary conditions necessary to determine the diffracted amplitudes outside of the grating.

These rigorous theories have been shown to reduce to the approximate theories: 1) two-wave modal theory, 2) two-wave second-order coupled-wave theory, 3) multiwave first-order coupled-wave theory, 4) two-wave first-order coupled-wave theory (Kogelnik theory), 5) Raman-Nath theory, and 6) amplitude transmittance theory in the appropriate limits. The assumptions associated with each of these approximate theories have been explicitly presented.

The rigorous theories presented in this paper (and thus their approximate versions) are based on the Floquet theorem. As such, they require a truly periodic grating (an infinite number of periods). These theories may be applied in the angular limit as the slanted fringes of a general grating approach being parallel to the surface (ϕ approaches zero). However, for exact parallelism with the surface ($\phi = 0$), the grating is no longer strictly periodic and a continuum of solutions is possible depending on the number of periods, the starting conditions, and the ending conditions of the grating. This pure reflection grating case can be analyzed without approximation using a rigorous chain-matrix method of analysis. This is discussed in [58].

Acknowledgements. This work was sponsored by the National Science Foundation and by the Joint Services Electronics Program.

References

1. R.R. Aggarwal: Proc. Indian Acad. Sci. **31**, 417–426 (1950)
2. P. Phariseau: Proc. Indiana Acad. Sci. A**44**, 165–170 (1956)
3. W.R. Klein, B.D. Cook: IEEE Trans. SU**14**, 123–134 (1967)
4. H. Kogelnik: Bell Syst. Tech. J. **48**, 2909–2947 (1969)
5. G.L. Fillmore, R.F. Tynan: J. Opt. Soc. Am. **61**, 199–203 (1971)
6. J.A. Kong: J. Opt. Soc. Am. **67**, 825–829 (1977)
7. R. Magnusson, T.K. Gaylord: J. Opt. Soc. Am. **67**, 1165–1170 (1977)
8. M.G. Moharam, T.K. Gaylord: J. Opt. Soc. Am. **71**, 811–818 (1981)
9. T. Tamir, H.C. Wang, A.A. Oliner: IEEE Trans. MTT-**12**, 323–335 (1964)
10. T. Tamir, H.C. Wang: Can. J. Phys. **44**, 2073–2094 (1966)
11. T. Tamir: Can. J. Phys. **44**, 2461–2494 (1966)
12. C.B. Burckhardt: J. Opt. Soc. Am. **56**, 1502–1509 (1966)
13. L. Bergstein, D. Kermisch: Proc. Symp. Mod. Opt. **17**, 655–680 (1967)
14. R.S. Chu, T. Tamir: IEEE Trans. MTT-**18**, 486–504 (1970)
15. R.S. Chu, T. Tamir: Proc. IEE **119**, 797–806 (1972)
16. F.G. Kaspar: J. Opt. Soc. Am. **63**, 37–45 (1973)
17. S.T. Peng, T. Tamir, H. L. Bertoni: IEEE Trans. MTT-**23**, 123–133 (1975)
18. R.S. Chu, J.A. Kong: IEEE Trans. MTT-**25**, 18–24 (1977)
19. P.St.J. Russell: Phys. Rep. **71**, 209–312 (1981)
20. L.Solymar, D.J. Cooke: *Volume Holography and Volume Gratings* (Academic Press, London 1981)
21. R. Magnusson, T.K. Gaylord: J. Opt. Soc. Am. **68**, 1777–1779 (1978)
22. D.A. Watkins: *Topics in Electromagnetic Theory* (Wiley, New York 1958) p. 2
23. R.E. Collin: *Field Theory of Guided Waves* (McGraw-Hill, New York 1960) p. 368
24. E.g.C.L. Liu, J.W.S. Liu: *Linear Systems Analysis* (McGraw-Hill, New York 1975)
25. E.g. program EIGRF from the International Mathematics and Statistics Library (IMSL)
26. F. Lederer, U. Langbein: Opt. Acta **27**, 183–200 (1980)
27. S. Kessler, R. Kowarschik: Opt. Quantum Electron. **7**, 1–4 (1975)
28. R. Kowarschik, S. Kessler: Opt. Quantum Electron. **7**, 399–411 (1975)
29. R. Kowarschik: Opt. Acta **23**, 1039–1051 (1976)
30. T. Jaaskelainen, O. Salminen, R.M.K. Hamalainen: Acta Poly.-Appl. Phys. **126**, 3–12 (1979)
31. C.V. Raman, N.S.N. Nath: Proc. Indian Acad. Sci. A**3**, 119–125 (1936)
32. C.V. Raman, N.S.N. Nath: Proc. Indian Acad. Sci. A**2**, 406–412 (1935)
33. C.V. Raman, N.S.N. Nath: Proc. Indian Acad. Sci. A**2**, 413–420 (1935)
34. C.V. Raman, N.S.N. Nath: Proc. Indian Acad. Sci. A**3**, 75–84 (1936)
35. C.V. Raman, N.S.N. Nath: Proc. Indian Acad. Sci. A**3**, 459–465 (1936)
36. N.S.N. Nath: Proc. Indian Acad. Sci. A**8**, 499–503 (1938)
37. N.S.N. Nath: Proc. Indian Acad. Sci. A**4**, 222–242 (1936)
38. M.V. Berry: *The Diffraction of Light by Ultrasound* (Academic Press, London 1966)
39. P.St.J. Russell, L. Solymar, M.P. Jordan: Proc. ICO 11, 635–638 (1978)
40. M.G. Moharam, T.K. Gaylord, R. Magnusson: J. Opt. Soc. Am. **70**, 300–304 (1980)
41. L. Solymar, M.P. Jordan: Microwave Opt. Acoust. **1**, 89–92 (1977)
42. R.P. Kenan: IEEE J. QE-**14**, 924–930 (1978)
43. M.G. Moharam, T.K. Gaylord, R. Magnusson: J. Opt. Soc. Am. **70**, 437–442 (1980)
44. N. Uchida: J. Opt. Soc. Am. **63**, 280–287 (1973)
45. U. Killat: Opt. Commun. **21**, 110–111 (1977)
46. M.P. Owen, L. Solymar: Opt. Commun. **34**, 321–326 (1980)

47. M.G. Moharam, T.K. Gaylord, R. Magnusson: Opt. Commun. **32**, 19–23 (1980)
48. F.H. Sanders: Can. J. Phys. **14**, 158–171 (1936)
49. M.G. Moharam, T.K. Gaylord, R. Magnusson: Opt. Commun. **32**, 14–18 (1980)
50. K.L. Zankel, E.A. Hiedemann: J. Acoust. Soc. Am. **31**, 44–54 (1959)
51. R. Magnusson, T.K. Gaylord: J. Opt. Soc. Am. **68**, 806–809 (1978)
52. R. Magnusson, T.K. Gaylord: Opt. Commun. **25**, 129–132 (1978).
53. J.W. Goodman: *Introduction to Fourier Optics* (McGraw-Hill, New York 1968)
54. W.T. Cathey: *Optical Information Processing and Holography* (Wiley, New York 1974)
55. H.M. Smith: *Principles of Holography* (Wiley, New York 1969)
56. R.J. Collier, C.B. Burckhardt, L.H. Lin: *Optical Holography* (Academic Press, New York 1971)
57. R. Magnusson, T.K. Gaylord: Opt. Commun. **28**, 1–3 (1979)
58. M.G. Moharam, T.K. Gaylord: J. Opt. Soc. Am. **72**, 187–190 (1982)

Reprinted from **APPLIED OPTICS**, Vol. 20, page 3271, October 1, 1981

Thin and thick gratings: terminology clarification

T. K. Gaylord and M. G. Moharam

Georgia Institute of Technology, School of Electrical Engineering, Atlanta, Georgia 30332.
Received 22 May 1981.
0003-6935/81/193271-03$00.50/0.

The terminology *thin* and *thick* gratings is widely used. However, these terms frequently either are not defined, vaguely defined, or defined in an ambiguous way. Interpretations of *thin* and *thick* grating behavior appear in the literature dating back to the 1930's. These interpretations with their various degrees of preciseness and accuracy have been carried forward in parallel in many cases. The terminology of *thin* and *thick* gratings is often confusing to workers in fields that use planar gratings (such as acoustooptics, holography, integrated optics, and spectroscopy). The purpose of this short paper is to clarify the possible practical explicit definitions of *thin* and *thick* gratings. This is done in terms of the diffraction regime characteristics and angular and wavelength selectivity characteristics of the grating. For brevity, only the common case of planar gratings with grating fringes perpendicular to the surfaces is discussed. However, planar absorption gratings and slanted fringe gratings may be similarly analyzed.

For a plane wave incident upon a planar grating with (co)-sinusoidal permittivity fringes normal to the surface, the fields inside the grating may be completely described by the rigorous coupled-wave equations.[1] These equations are obtained by substituting the periodic relative permittivity, the plane wave field expansion, and the Floquet theorem into the wave equation. For H-mode polarization the result is

$$\frac{1}{2\pi^2}\frac{d^2S_i(z)}{dz^2} - j\frac{2(\epsilon_0)^{1/2}\cos\theta}{\pi\lambda}\frac{dS_i(z)}{dz} + \frac{2i(m-i)}{\Lambda^2}S_i(z) + \frac{\epsilon_1}{\lambda^2}[S_{i+1}(z) + S_{i-1}(z)] = 0, \quad (1)$$

where $S_i(z)$ are the fields inside the grating, i is the (integer) order of diffraction, $(-\infty < i < +\infty)$, ϵ_0 is the average relative permittivity, ϵ_1 is the amplitude of the (co)sinusoidal relative permittivity, λ is the free space wavelength, θ is the angle of refraction of the incident wave, Λ is the grating period, and m is given by

$$m\lambda/(\epsilon_0)^{1/2} = 2\Lambda\sin\theta. \quad (2)$$

This is a Bragg condition when m is an integer. Equation (1) is derived without approximation.

The Raman-Nath diffraction regime[2] may be obtained from Eq. (1) by neglecting the d^2S_i/dz^2 and S_i terms. The Bragg diffraction regime may be obtained from Eq. (1) by neglecting the d^2S_i/dz^2 term and considering only $i = 0,1$. This corresponds to the Kogelnik two-wave coupled-wave theory.[3] In any case, the diffraction efficiency is given by $\eta_i = S_iS_i^*$ for unslanted phase gratings.

A *thin* grating may be described as a grating that produces Raman-Nath regime diffraction.[2] In this case, the multiple grating diffracted-orders ideally have diffraction efficiencies η_i given by

$$\eta_i = J_i^2(2\gamma), \quad (3)$$

where i is the integer representing the diffracted-order, J_i is an integer-order ordinary Bessel function of the first kind, and γ is the grating strength parameter given by $\gamma = \pi\epsilon_1 d/2\lambda(\epsilon_0)^{1/2}\cos\theta$, and d is the grating thickness. The occurrence (or lack of occurrence) of Raman-Nath regime diffraction as given by Eq. (3) may be determined using a number of different criteria[4] depending on the application of the grating. The practical criteria and their interpretations are: (1) *Zero-Order Beam Criterion*—the transmitted (zero-order) diffraction efficiency is predicted by Eq. (3) to within some specified limit; (2) *First-Order Beam Criterion*—the fundamental (first-order) diffraction efficiency is predicted by Eq. (3) to within some specified limit; and (3) *Composite Criterion*—all diffracted-orders simultaneously have diffraction efficiencies predicted by Eq. (3) to within some specified limit. Each of these criteria is evaluated in Ref. 4. Those evaluations showed that each of these criteria is met to within 1% diffraction efficiency when the condition

$$Q'\gamma \le 1 \quad (4)$$

is satisfied. Q' is a grating parameter given by $Q' = Q/\cos\theta = 2\pi\lambda d/(\epsilon_0)^{1/2}\Lambda^2\cos\theta$. This condition was originally used by Extermann and Wannier[5] and later in the form of $Q'\gamma \le \pi^2/8$ by Willard.[6] From Eq. (4) it is apparent that Raman-Nath regime diffraction behavior will be observed for any value of γ (proportional to grating modulation ϵ_1) if Q' is sufficiently small. This has led to the incomplete popular condition of $Q' < 1$ for describing *thin* gratings.[7] The first-order diffraction efficiency $J_1^2(2\gamma)$ for ideal Raman-Nath regime diffraction is shown in Fig. 1.

A *thick* grating (or *volume* grating) may be described as a grating that produces Bragg regime (or two-wave regime) diffraction. This is described by the two-wave coupled-wave

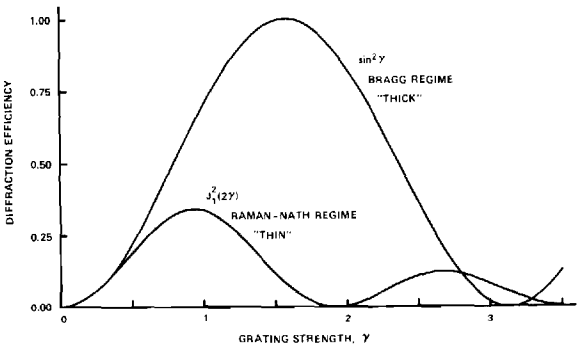


Fig. 1. Ideal first-order diffraction efficiencies for Raman-Nath (*thin* grating) regime [$\eta_1 = J_1^2(2\gamma)$] and for Bragg (*thick* grating) regime ($\eta_1 = \sin^2\gamma$).

theory of Kogelnik.[3] In this regime, the single fundamental diffracted-order ideally has a diffraction efficiency given by

$$\eta_1 = \sin^2\gamma. \tag{5}$$

Correspondingly, the transmitted (zero-order) wave has a diffraction efficiency given by $\cos^2\gamma$. For E-mode polarization the γ in Eq. (5) is $\gamma = \pi\epsilon_1 d \cos 2\theta/2\lambda(\epsilon_0)^{1/2}\cos\theta$. The occurrence (or lack of occurrence) of Bragg regime diffraction as given by Eq. (5) may be determined using a number of different criteria[8] depending on the application. The practical criteria include the *Zero-Order Beam Criterion* and *First-Order Beam Criterion* defined as before except that the deviation is with respect to Eq. (5). The other criteria and their interpretations are: *Two-Wave Criterion*—the sum of the diffraction efficiency associated with all higher-order waves is less than some specified limit; and *Composite Criterion*—all three criteria are met to within some specified limit. Each of these criteria is evaluated in Ref. 8. Those evaluations showed that each of these criteria is met to within 1% diffraction efficiency when the condition

$$\rho \equiv Q'/2\gamma \geq 10 \tag{6}$$

is satisfied. The parameter ρ was first used by Nath.[9] It is now clear that this parameter determines the Bragg regime diffraction boundary. Since $\rho = 2\lambda^2/\Lambda^2\epsilon_1$ for H mode and $\rho = 2\lambda^2/\Lambda^2\epsilon_1\cos 2\theta$ for E mode, it is particularly notable that this diffraction-regime-based definition of a *thick* grating is independent of grating thickness. If, in addition to the above criteria, it is also required that $\eta_1 > \eta_{-1}$ (for example, in holography, to make the power in the conjugate image beam be small compared with that in the primary diffracted beam), the condition $Q' > 1$ is also needed in addition to (6).[10,11] From Eq. (6) it is apparent that Bragg regime diffraction behavior will occur for any value of γ (or ϵ_1) if Q' is sufficiently large. Thus, the incomplete condition $Q' > 1$ is often used alone to describe *thick* gratings.[3,7]

For exact Bragg regime diffraction, the diffraction efficiency is given by $\sin^2\gamma$ and this is also shown in Fig. 1. For small values of the grating strength γ, the diffraction efficiencies of the fundamental diffracted-order converge to the same diffraction efficiency $\eta_1 = \gamma^2$ for both *thin* and *thick* gratings. When both conditions (4) and (6) are simultaneously satisfied the distinction between *thin* and *thick* is lost when using diffraction-regime-based definitions.

When neither condition (4) nor (6) is satisfied, the grating will be neither *thin* nor *thick* according to diffraction-regime-based definitions. The first-order diffraction efficiency will not be accurately predicted by either Eqs. (3) or (5). In general, in the intermediate regime, a rigorous (without approximations) planar grating diffraction theory (e.g., Refs. 1, 12, and 13) is required.

A *thin* grating may be alternatively described as a grating exhibiting relatively little angular and wavelength selectivity. As the incident wave is dephased (either in angle incidence or in wavelength) from the Bragg condition, the diffraction efficiency decreases. The angular range or wavelength range for which the diffraction efficiency decreases to half of its on-Bragg-angle value is determined by the thickness of the grating d expressed as a number of grating periods Λ. For a *thin* grating this number may be reasonably chosen to be

$$d/\Lambda < 10. \tag{7}$$

The region of *thin* grating behavior according to the angular-and-wavelength-selectivity-based definition (7) is depicted in Fig. 2. Gratings having angular and wavelength selectivities with FWHM wider than that for $d/\Lambda = 10$ may be considered to be *thin* gratings. This definition does not accurately predict the diffraction regime. It has the desirable feature that the governing parameter (d/Λ) is directly proportional to the grating thickness, and thus *thin* and *thick* have direct physical interpretations.

A *thick* grating may conversely be described as a grating exhibiting strong angular and wavelength selectivity. A relatively small change in the angle of incidence from the Bragg angle or a relatively small change in the wavelength at the Bragg angle produces significant dephasing and the diffraction efficiency decreases correspondingly. *Thick* grating behavior may be considered to occur when

$$d/\Lambda > 10. \tag{8}$$

This is the angular-and-wavelength-selectivity-based definition of a *thick* grating. The region of this behavior is also shown in Fig. 2. Unlike with the diffraction-regime-based definitions ($\gamma Q' \leq 1$ and $Q'/2\gamma \geq 10$), there is a single simple boundary ($d/\Lambda = 10$) between *thin* and *thick* grating diffraction.

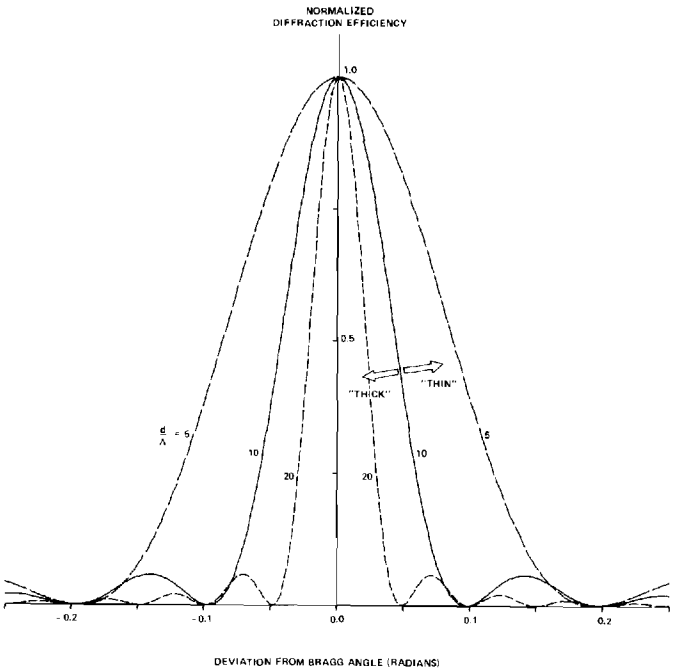


Fig. 2. Typical angular selectivity plots (normalized first-order diffraction efficiency vs angular deviation from the Bragg angle) for various values of d/Λ.

The terms *thin* and *thick* gratings may be ambiguous. However, distinct meaningful definitions are possible based on either the diffraction regime or on the angular and wavelength selectivity. These definitions may be concisely summarized as follows:

(1) If *thin* grating is intended to mean Raman-Nath regime diffraction, then the required condition is $Q'\gamma \leq 1$.

(2) If *thin* grating is intended to mean broad angular and wavelength selectivity, then the required condition is $d/\Lambda < 10$.

(3) If *thick* grating is intended to mean Bragg regime diffraction, then the required condition is $\rho \geq 10$. (If $\eta_1 > \eta_{-1}$ is also included in the definition of Bragg regime, $Q' > 1$ is required in addition.)

(4) If *thick* grating is intended to mean narrow angular and wavelength selectivity, then the required condition is $d/\Lambda > 10$.

This work was sponsored by the National Science Foundation and the Joint Services Electronics Program.

References

1. M. G. Moharam and T. K. Gaylord, J. Opt. Soc. Am. **71,** 811 (1981).
2. C. V. Raman and N. S. N. Nath, Proc. Indian Acad. Sci. **2,** 406, 413 (1935); **3,** 75, 119 (1936).
3. H. Kogelnik, Bell Syst. Tech. J. **48,** 2909 (1969).
4. M. G. Moharam, T. K. Gaylord, and R. Magnusson, Opt. Commun. **32,** 19 (1980).
5. R. Extermann and G. Wannier, Helv. Phys. Acta **9,** 520 (1936).
6. G. W. Willard, J. Acoust. Soc. Am. **21,** 101 (1949).
7. W. R. Klein and B. D. Cook, IEEE Trans. Sonics Ultrason. **SU-14,** 123 (1967).
8. M. G. Moharam, T. K. Gaylord, and R. Magnusson, Opt. Commun. **32,** 14 (1980).
9. N. S. N. Nath, Proc. Indian Acad. Sci. **8,** 499 (1938).
10. P. Phariseau, Proc. Indian Acad. Sci. Sect. A **44,** 165 (1965).
11. B. Benlarhi, D. J. Cooke, and L. Solymar, Opt. Acta. **27,** 885 (1980).
12. C. B. Burckhardt, J. Opt. Soc. Am. **56,** 1502 (1966).
13. R. S. Chu and T. Tamir, IEEE Trans. Microwave Theory Tech. **MTT-18,** 486 (1970).

E-21-659

FINAL REPORT

RESEARCH INITIATION:
A STUDY OF PLANAR DIELECTRIC GRATING COUPLERS

by

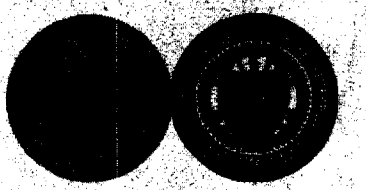
M. G. Moharam

Performed for

NATIONAL SCIENCE FOUNDATION
Division of Electrical, Computer, and System Engineering

Under Grant No. ESC-8105483
1 July 1981 to 31 December 1983

GEORGIA INSTITUTE OF TECHNOLOGY
A UNIT OF THE UNIVERSITY SYSTEM OF GEORGIA
SCHOOL OF ELECTRICAL ENGINEERING
ATLANTA, GEORGIA 30332



APPENDIX VII

NATIONAL SCIENCE FOUNDATION
Washington, D.C. 20550

FINAL PROJECT REPORT
NSF FORM 98A

PLEASE READ INSTRUCTIONS ON REVERSE BEFORE COMPLETING

PART I–PROJECT IDENTIFICATION INFORMATION

1. Institution and Address	2. NSF Program	3. NSF Award Number
Georgia Institute of Technology Atlanta, Georgia 30332	Research Initiation	ECS-8105483
	4. Award Period From 7/1/81 To 12/31/83	5. Cumulative Award Amount $48,000

6. Project Title

"Research Initiation: A Study of Planar Dielectric Grating Couplers"

PART II–SUMMARY OF COMPLETED PROJECT *(FOR PUBLIC USE)*

A program of study of planar dielectric grating couplers was undertaken. The study focused on the problems associated with using dielectric grating couplers to achieve high efficiency beam coupling to planar waveguide structures.

A number of important problems were identified and significant analytical results were obtained in the study of these areas. A new and exact rigorous electromagnetic coupled-wave theory to analyze grating diffraction has been developed. This new theory allows exact prediction of the diffraction characteristics of these periodic structures. The dependence on the various parameters has been extensively and systematically studied. The theory is applicable to both TE and TM polarization. The coupled-wave approach was extended to surface-relief dielectric grating thus enabling the study of surface-relief dielectric grating couplers. A three-dimensional rigorous coupled-wave theory has been developed to analyze gratings where the direction of the periodicity is not in the plane of incidence, thus resulting in conical diffraction.

Knowledge of the exact diffraction characteristics of diffraction grating will result in significant improvement in the design and the fabrication of devices that utilize a diffraction grating in its operation (for example grating couplers, acoustic and electro optic modulators, grating lens, and holographic optical elements.)

PART III–TECHNICAL INFORMATION *(FOR PROGRAM MANAGEMENT USES)*

1. ITEM *(Check appropriate blocks)*	NONE	ATTACHED	PREVIOUSLY FURNISHED	TO BE FURNISHED SEPARATELY TO PROGRAM	
				Check (√)	Approx. Date
a. Abstracts of Theses		X			
b. Publication Citations		X			
c. Data on Scientific Collaborators					
d. Information on Inventions	X				
e. Technical Description of Project and Results		X			
f. Other *(specify)*					

2. Principal Investigator/Project Director Name *(Typed)*	3. Principal Investigator/Project Director Signature	4. Date
M.G. Moharam		

NSF Form 98A (3-83) Supersedes All Previous Editions

Form Approved OMB No. 3145-0058

FINAL REPORT

RESEARCH INITIATION:

A STUDY OF PLANAR DIELECTRIC GRATING COUPLERS

by

M. G. Moharam
School of Electrical Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332

Performed for

NATIONAL SCIENCE FOUNDATION
Division of Electrical, Computer, and System Engineering
Grant No. ESC-8105483
1 July 1981 to 31 December 1983

ABSTRACT

A program of study of planar dielectric grating couplers was undertaken. The study focused on the problems associated with using dielectric grating couplers to achieve high efficiency beam coupling to planar waveguide structures.

A number of important problems were identified and significant analytical results were obtained in the study of these areas. A new and exact rigorous electromagnetic coupled-wave theory to analyze grating diffraction has been developed. This new theory allows the exact prediction of the diffraction characteristics of these periodic structures. The dependence on the various parameters has been extensively and systematically studied. The theory is applicable to both TE and TM polarization. The coupled-wave approach was extended to surface-relief dielectric grating thus enabling the study of surface-relief dielectric grating couplers. A three-dimensional rigorous coupled-wave theory has been developed to analyze gratings where the direction of the periodicity is not in the plane of incidence, thus resulting in conical diffraction.

Knowledge of the exact diffraction characteristics of diffraciton grating will result in significant improvements in the design and the fabrication of devices that utilize diffraction grating in their operation (for example, grating couplers, acoustic and electro-optic modulators, grating lens, and holographic optical elements).

DESCRIPTION OF PROJECT AND RESULTS

The purpose of this project is to study dielectric planar grating couplers and to determine its coupling characteristics when integrated into a planar dielectric waveguide strucure. The study was to investigate the coupling efficiency and its dependence on the various grating and waveguide paramenters.

The first phase of this research was to investigate the diffraction characteristic of planar dielectric gratings. Existing approximate theories are not applicable to structures where the refraction and transmission at the boundaries are essential to the analysis, such structures include grating couplers on dielectric planar waveguides. In addition, approximate theories are limited in applicability to situations where the approximates can be tolerated, which may not be the case in grating couplers where maximizing the coupling efficiency depends on both the grating and the dielectric waveguide.

To overcome the limitations, a new exact electromagnetic theory of grating diffraction was needed. Such a theory, the rigorous coupled-wave theory has been developed [1-2]. For the first the exact diffraction characteristics of planar dielectric gratings have been calculated without approximations and without limitations on the range of the different grating parameters. The rigorous coupled-wave analysis which is applicable to both TE and TM polarizations has been used for an extensive study of the dependence of the diffraction characteristics on the grating parameters such as grating period, grating slant, permittivity modulation, incident light wavelength, angle of incidence, and grating thickness. The theory has been also used to evaluate the existing approximate theories and to study the effect of the many approximations on the accuracy of the predicted diffraction characteristics [1-3].

This exact rigorous coupled-wave theory has been extended to surface-relief dielectric gratings [4] and for the first time the diffraction characteristics of surface-relief gratings of various profiles are determined. It is shown for the first time that transmission dielectric surface-relief gratings can diffract almost 100% of the incident energy into one diffracted order [4]. This new finding generated great interest in surface-relief gratings and they are considered by many as superior to conventional holographic planar gratings [5].

As a secondary result the rigorous coupled-wave theory is applied to reflection gratings [6,7] which play an important role in distributed-feedback semi-conductor lasers. It is also applied to pure absorption gratings [8] where the maximum possible diffraction efficiency of these gratings has been rigorously determined. In the design of grating devices an important consideration is the regime of diffraction, that is, two wave or Bragg regime an multi-wave or Raman-Nath regime of diffraction. Using the exact coupled-wave theory, a better criteria for the two regimes have been developed [8].

In some recent grating devices, the grating vector (the direction of periodicity) is not in the plane of incidence. In this configuration the diffracted waves directions are on the surface of a cone. In this conical diffraction, as opposed to planar diffraction, the electromagnetic fields may not be decoupled into TE and TM polarizations. An exact rigorous three-dimensional coupled-wave theory has been developed to analyze conical diffraction [10]. An extensive study of the angular selectivity of such gratings has been carried out with interesting results and conclusions [10].

The calculated exact diffraction characteristics of square wave dielectric gratings on a planar waveguide structure have been compared to the experimental results supplied by Professor W. S. C. Chang of the University

of California at San Diego [11]. Excellent agreement has been obtained indicating that the coupled-wave theory is a powerful, accurate, and useful technique to analyze grating diffraction.

As a result of the expertise developed in the grating diffraction, an extensive review paper on diffraction gratings and their applications is to be published as an invited paper in the Proceedings of IEEE [12].

SUMMARY

Significant contributions have been made on the understanding and the analysis of grating diffraction. An exact rigorous coupled-wave theory to analyze grating diffraction has been developed. This unified theory is applicable without approximations or limitations to planar and surface-relief grating and to dielectric (phase) and metallic (absorption) gratings. The theory is applicable to configurations where the electromagnetic fields may not be decoupled into TE and TM polarizations. This method of analysis is having and will have a significant impact on the analysis and the design of grating devices.

Based on the work performed under this grant, over 12 refereed papers, and over 7 abstracts have been published, and over 9 presentations have been made.

PUBLICATIONS

[1] Moharam, M. G. and Gaylord, T. K., "Rigorous coupled-wave analysis of planar gratings," J. Opt. Soc. Am., Vol. 71, pp. 811-818, July 1981.

[2] Moharam, M. G. and Gaylord, T. K., "Rigorous coupled-wave analysis of grating diffraction--E mode polarization and losses," J. Opt. Soc. Am., Vol. 73, pp. 451-455, July 1983.

[3] Gaylord, T. K. and Moharam, M. G., "Planar grating diffraction theories," Applied Physics B, Vol. 27, pp. 1-14, May 1982. (invited)

[4] Moharam, M. G. and Gaylord, T. K., "Diffraction analysis of dielectric surface-relief gratings," J. Opt. Soc. Am., Vol. 72, pp. 1385-1392, October 1982.

[5] Moharam, M. G., Gaylord, T. K., Sincerbox, G. T., Werlich, H., and Yung, B., "Diffraction characteristics of photoresist surface-relief gratings," Applied Optics, Vol. 23, pp. 3214-3220, September 1984.

[6] Moharam, M. G. and Gaylord, T. K., "Rigorous chain matrix analysis of arbitrary-thickness dielectric reflection gratings," J. Opt. Soc. Am., Vol. 72, pp. 187-190, February 1982.

[7] Moharam, M. G. and Gaylord, T. K., "Comments on analyses of reflection gratings," J. Opt. Soc. Am., Vol. 72, pp. 399-401, March 1983.

[8] Baird, W. E., Moharam, M. G., and Gaylord, T. K., "Diffraction characteristics of planar absorption gratings," Applied Physics B, Vol. 32, pp. 15-20, 1983.

[9] Gaylord, T. K. and Moharam, M. G., "Thin and thick gratings: terminology clarification," Applied Optics, Vol. 20, pp. 3271-3273, October 1981.

[10] Moharam, M. G. and Gaylord, T. K., "Three-dimensional vector coupled-wave analysis of planar grating diffraction," J. Opt. Soc. Am., Vol. 73, pp. 1105-1112, September 1983.

[11] Delavaux, J. M., Chang, W. S. C., and Moharam, M. G., "Comparison of the experimental and theoretical diffraction characteristics of transmission gratings on planar dielectric waveguides," Applied Optics, Vol. 23, pp. 0000-0000, December 1984.

[12] Gaylord, T. K. and Moharam, M. G., "Analysis and applications of grating diffraction," Proc. IEEE, Vol. 72, pp. 0000-0000, 1985. (invited)

ABSTRACTS AND PRESENTATIONS

[13] Moharam, M. G. and Gaylord, T. K., "Rigorous coupled-wave analysis of surface grating with arbitrary profiles," J. Opt. Soc. Am., Vol. 71, p. 1573, December 1981 (Abstract). Also presented at the Optical Society of America Annual Meeting, Orlando, Florida, October 1981.

[14] Gaylord, T. K. and Moharam, M. G., "Interrelationship between various planar grating diffraction theories," J. Opt. Soc. Am. Vol. 71, p. 1569, December 1981 (Abstract). Also presented at the Optical Society of America Annual Meeting, Orlando, Florida, October 1981.

[15] Moharam, M. G. and Gaylord, T. K., "Vector three-dimensional theory for planar grating diffraction," J. Opt. Soc. Am., Vol. 72, p. 1813, December 1982 (Abstract). Also presented at the Optical Society of America Annual Meeting, Tucson, Arizona, October 1982.

[16] Moharam, M. G. and Gaylord, T. K., "Rigorous theory of planar-reflection-grating diffraction," J. Opt. Soc. Am., Vol. 72, p. 1824, December 1982 (Abstract). Also presented at the Optical Society of America Annual Meeting, Tucson, Arizona, October 1982.

[17] Baird, W. E., Gaylord, T. K., and Moharam, M. G., "Diffraction efficiencies of transmission absorption gratings," J. Opt. Soc. Am., Vol. 73, p. 1889, December 1983 (Abstract). Also presented at the Optical Society of America Annual Meeting, New Orleans, Louisiana, October 1983.

[18] Moharam, M. G. and Gaylord, T. K., "Diffraction of finite beams by dielectric gratings," J. Opt. Soc. Am., Vol. 73, p. 1941, December 1984 (Abstract). Also presented at the Optical Society of American Annual Meeting, New Orleans, Louisiana, October 1983.

[19] Moharam, M. G., Gaylord, T. K., Sincerbox, G. T., Werlich, H., and Yung, B., "Diffraction characteristics of surface-relief gratings," J. Opt. Soc. Am., Vol. 73, p. 1941, December 1981 (Abstract). Also presented at the Optical Society of America Annual Meeting, New Orleans, Louisiana, October 1983.

Rigorous coupled-wave analysis of planar-grating diffraction

M. G. Moharam and T. K. Gaylord

School of Electrical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

Received July 7, 1980; revised manuscript received February 18, 1981

A rigorous coupled-wave approach is used to analyze diffraction by general planar gratings bounded by two different media. The grating fringes may have any orientation (slanted or unslanted) with respect to the grating surfaces. The analysis is based on a state-variables representation and results in a unifying, easily computer-implementable matrix formulation of the general planar-grating diffraction problem. Accurate diffraction characteristics are presented for the first time to the authors' knowledge for general slanted gratings. This present rigorous formulation is compared with rigorous modal theory, approximate two-wave modal theory, approximate multiwave coupled-wave theory, and approximate two-wave coupled-wave theory. Typical errors in the diffraction characteristics introduced by these various approximate theories are evaluated for transmission, slanted, and reflection gratings. Inclusion of higher-order waves in a theory is important for obtaining accurate predictions when forward-diffracted orders are dominant (transmission-grating behavior). Conversely, when backward-diffracted orders dominate (reflection-grating behavior), second derivatives of the field amplitudes and boundary diffraction need to be included to produce accurate results.

1. INTRODUCTION

The diffraction of electromagnetic waves by planar gratings has been extensively studied in recent years. These periodic structures have applications in several diverse areas, such as acousto-optics, integrated optics, holography, and spectroscopy. Several different techniques have been used to analyze the diffraction of electromagnetic waves by spatially modulated media. The most common of these methods are the coupled-wave approach[1-7] and the modal approach.[8-17] Both of these approaches are capable of being formulated in a rigorous manner without approximations. As such they have been shown to be exact and equivalent.[18] However, until the present work, accurate calculations have been rather difficult and time consuming. This paper presents a straightforward rigorous coupled-wave analysis that is easily implemented on a computer. The method of solution is based on a state-space representation of the governing set of differential equations.

The coupled-wave approach has long been known to offer superior physical insight into wave-diffraction phenomena, and it frequently yields simple analytical results in limiting cases. In this approach several assumptions typically are made in order to obtain solutions. These approximations are (1) neglecting boundary diffraction,[1-5,7] (2) neglecting the second derivatives of the field amplitudes,[1-5,7] and (3) retaining only one diffracted wave[1,2,4,6] (in addition to the transmitted wave). The first two of these approximations are interconnected and cannot be separated for the case of a planar grating. The full modal approach is a rigorous exact analysis, but solutions have been published only for the unslanted pure transmission case in which the grating vector is parallel to the grating surface. Several authors have attempted to analyze general slanted gratings. Kogelnik[4] and Magnusson and Gaylord[7] applied the coupled-wave approach, with various assumptions discussed above, to this problem. Clearly these analyses are not exact, and in some cases they give obviously incorrect results because of the use of these approximations. Bergstein and Kermisch[12] applied the modal approach to slanted gratings, but they assumed a two-wave regime to obtain solutions. Chu and Kong[17] have also presented a generalized modal approach formulation for the general slanted-gratings case. However, they did not present any calculated results for general slanted gratings because the formulation for slanted gratings in the modal approach results in a complicated transcendental relationship that may be difficult to solve in general. This paper presents, for the first time to the authors' knowledge, accurate diffraction results for slanted gratings.

By using the straightforward rigorous method of analysis presented in this paper, it is a simple matter to analyze the precise effect of using the various approximations mentioned above. The effect of neglecting boundary diffraction and second derivatives of field amplitudes is studied by comparing the present theory with multiwave coupled-wave theory.[7] The effect of neglecting higher-order waves (retaining only one diffracted wave) is analyzed by comparing the present theory and the two-wave modal theory.[12] The combined effect of these approximations is studied by comparing the present theory with the two-wave coupled-wave theory.[4] These three errors are calculated and presented for the first time for several typical transmission, reflection, and general planar gratings.

The rigorous coupled-wave approach presented here analyzes the diffraction of an electromagnetic plane wave incident obliquely at a planar grating bounded by two different media. In general, these gratings simultaneously exhibit both transmission and reflection behavior, as indicated by the forward-diffracted and backward-diffracted waves shown in Fig. 1. This analysis is applicable (1) to holographic gratings in air or other media ($\epsilon_1 = \epsilon_3 \neq \epsilon_2$), (2) to acousto-optic gratings within a medium ($\epsilon_1 = \epsilon_3 = \epsilon_2$), and (3) to grating couplers such as are used in integrated optics ($\epsilon_1 \neq \epsilon_2 \neq \epsilon_3 \neq \epsilon_1$).

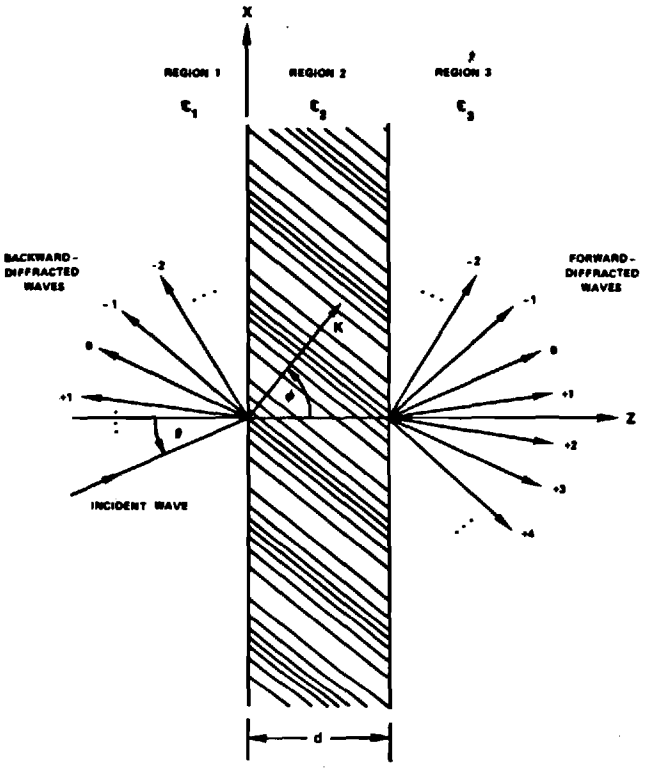


Fig. 1. Geometry for planar-grating diffraction.

2. THEORY

As in most of the work previously cited, the problem under consideration is the diffraction of an obliquely incident plane wave on a lossless (pure phase) sinusoidal grating with the incident wave polarized perpendicular to the plane of incidence (H mode). Therefore the electric field will have only one component (in the y direction of Fig. 1). The relative permittivity in the modulated region ($0 < z < d$) is

$$\epsilon(x,z) = \epsilon_2 + \Delta\epsilon \cos[K(x \sin\phi + z \cos\phi)], \quad (1)$$

where ϵ_2 is the average dielectric constant, $\Delta\epsilon$ is the amplitude of the sinusoidal relative permittivity, ϕ is the grating slant angle, and $K = 2\pi/\Lambda$, where Λ is the grating period. The dielectric constant in the unmodulated regions ($z < 0$ and $z > d$) is ϵ_1 and ϵ_3, respectively. It is assumed that each of the three regions has the permeability of free space.

The general approach to the planar-grating problem involves finding a solution of the wave equation in each of the three regions and then matching the tangential electric and magnetic fields at the two interfaces ($z = 0$ and $z = d$) to determine the unknown constants (resulting from solving the differential-wave equation). In region 1, backward-diffracted waves exist. In general, these waves are produced both by diffraction from within the grating volume (bulk diffraction) and by diffraction from the periodic boundary at $z = 0$ (boundary diffraction). These diffraction processes produce a spectrum of plane waves traveling back into region 1 ($z < 0$). The normalized wave amplitudes in region 1 may be expressed as[8-10]

$$E_1 = \exp[-j(\beta_0 x + \xi_{10} z)] + \sum_i R_i \exp[-j(\beta_i x - \xi_{1i} z)], \quad (2)$$

where $\beta_i = k_1 \sin\theta - iK \sin\phi$ for any integer i (the wave index); $\xi_{li}^2 = k_l^2 - \beta_i^2$ for $l = 1, 3$ (the region index); $k_l = 2\pi\epsilon_l^{1/2}/\lambda$ for $l = 1, 2, 3$; λ is the free-space wavelength; $j = (-1)^{1/2}$; θ is the angle of incidence in region 1; and R_i is the normalized amplitude of the ith reflected wave and is to be determined from the matching of the electric and magnetic fields. In region 3 ($z > d$) the spectrum of transmitted plane waves may be expressed as[16,17]

$$E_3 = \sum_i T_i \exp\{-j[\beta_i x + \xi_{3i}(z - d)]\}, \quad (3)$$

where T_i is the normalized amplitude of the ith transmitted wave to be determined from the field matching. In region 2, the modulated region ($0 < z < d$), the electric field may be expressed as

$$E_2 = \sum_i \hat{S}_i(z) \exp[-j(\beta_i x + \xi_{2i} z)], \quad (4)$$

where $\xi_{2i} = k_2 \cos\theta' - iK \cos\phi$, θ' is the angle of refraction inside the modulated region, and $\hat{S}_i(z)$ is the normalized amplitude of the ith wave field at any point within the modulated region. For a given value of i, the wave field inside the grating is not a simple plane wave. It may be expressed as a superposition of an infinite number of plane waves (inherent in the coupled-wave formulation). This superposition includes forward-traveling waves (components in $+z$ direction) and corresponding backward-traveling waves (components in $-z$ direction). These amplitudes are to be determined from solving the modulated-region wave equation

$$\nabla^2 E_2 + (2\pi/\lambda)^2 \epsilon(x,z) E_2 = 0. \quad (5)$$

Note that the three electric fields E_1, E_2, and E_3 in the three regions are phase matched along the two interfaces. It is important to point out that the above analysis is valid for all slant angles ϕ except when the slant angle is identically zero (pure reflection grating). In this case, the modulation is no longer periodic since it has only a finite number of cycles. Equations (2)–(4) are thus not valid because they are derived from the Floquet theorem, which is correct only for infinite periodic structures. However, the above analysis can be used to analyze pure reflection gratings ($\phi = 0$) by considering $\phi = \delta$ and allowing δ to become arbitrarily small. As ϕ approaches (but does not reach) zero, the slab retains a periodicity in the modulation along its boundaries. All the R_i and T_i amplitudes for the reflected and transmitted waves may then be calculated. As ϕ approaches zero, all the reflected wave vectors in region 1 approach a common direction, and the resulting reflected intensity of the composite wave is $\Sigma_i |R_i|^2$. Likewise, all the transmitted wave vectors in region 3 approach a single direction, and the resulting transmitted intensity of this composite wave is $\Sigma_i |T_i|^2$. For a more detailed discussion of planar reflection gratings, see Ref. 19.

3. METHOD OF SOLUTION

To obtain the diffracted amplitudes $\hat{S}_i(z)$, Eqs. (1) and (4) are substituted into Eq. (5), resulting in the infinite set of coupled-wave equations

$$(\Delta\epsilon/8\epsilon_2)\frac{d^2 S_i(u)}{du^2} = (\cos\theta' - i\mu\cos\phi) \times \frac{dS_i(u)}{du} - \rho i(i - B)S_i(u) + S_{i+1}(u) + S_{i-1}(u), \quad (6)$$

with $S_i(u) = S_i(z)$, $u = j\pi\Delta\epsilon z/2\lambda(\epsilon_2)^{1/2} = j\kappa z$, $\mu = \lambda/\Lambda(\epsilon_2)^{1/2}$, $\rho = 2\lambda^2/\Lambda^2\Delta\epsilon = 2\mu^2\epsilon_2/\Delta\epsilon$, and $B = 2\Lambda(\epsilon_2)^{1/2}\cos(\phi - \theta')/\lambda = 2\cos(\phi - \theta')/\mu$. This system of coupled-wave equations has been derived *without* the common assumptions and approximations associated with previous coupled-wave analyses,[1-7] such as neglecting the second derivatives of the amplitudes. Therefore the present analysis is as rigorous and as exact as the modal approach. There are five fundamental parameters $[\theta', \phi, \Delta\epsilon/\epsilon_2, d/\Lambda, \lambda/\Lambda(\epsilon_2)^{1/2}]$ contained within Eq. (6). The composite parameters used in Eq. (6) may be expressed in terms of these parameters. Other equivalent sets of parameters obviously may also be chosen. The regime parameter ρ determines the boundary between the Bragg regime and the intermediate diffraction regime.[20] The parameter B represents the Bragg condition, i.e. for incidence at the pth Bragg condition $B = p$ (note that B is not an integer for off-Bragg incidence). The parameter κ is the widely used coupling coefficient. As is shown in the Appendix, if S and S' are defined as state variables, Eq. (6) may be written in matrix form as

$$\begin{bmatrix} \mathbf{S}' \\ \mathbf{S}'' \end{bmatrix} = \begin{bmatrix} b_{rs} \end{bmatrix} \begin{bmatrix} \mathbf{S} \\ \mathbf{S}' \end{bmatrix}, \tag{7}$$

where $\mathbf{S}$, $\mathbf{S}'$, and $\mathbf{S}''$ indicate the column vectors of S_i, dS_i/du, and d^2S_i/du^2, respectively. The quantity $[b]$ is the coefficient matrix specified from Eq. (6). Equation (7) corresponds to an unforced state equation in the state-space description of linear systems. The system of differential equations given by Eq. (7) has a relatively simple and straightforward solution, obtainable in terms of the eigenvalues and the eigenvectors of the coefficient matrix $[b]$. It is

$$S_i(u) = \sum_m C_m w_{im} \exp(q_m u), \tag{8}$$

where q_m is the mth eigenvalue and w_{im} is the mth element of the row in the matrix $[w]$ composed of the eigenvectors, corresponding to the ith wave (not the ith row of the matrix $[w]$). The coefficients C_m are unknown constants to be determined together with R_i and T_i by matching the tangential electric and magnetic fields at the two boundaries ($z = 0$ and $z = d$). The four quantities to be matched and the resulting boundary conditions are tangential E at $z = 0$:

$$R_i + \delta_{i0} = \sum_m C_m w_{im}, \tag{9}$$

tangential H at $z = 0$:

$$\xi_{1i}(R_i - \delta_{i0}) = \sum_m C_m w_{im}(q_m\kappa - \xi_{2i}), \tag{10}$$

tangential E at $z = d$:

$$T_i = \sum_m C_m w_{im} \exp[j(q_m\kappa - \xi_{2i})d], \tag{11}$$

tangential H at $z = d$:

$$-\xi_{3i}T_i = \sum_m C_m w_{im}(q_m\kappa - \xi_{2i}) \exp[j(q_m\kappa - \xi_{2i})d], \tag{12}$$

where δ_{i0} is the Kronecker delta function. Eliminating T_i and R_i from these equations gives

$$2\delta_{i0}\xi_{1i} = \sum_m C_m w_{im}(\xi_{2i} - q_m\kappa + \xi_{1i}), \tag{13}$$

$$0 = \sum_m C_m w_{im}(\xi_{2i} - q_m\kappa - \xi_{3i}) \exp(jq_m\kappa). \tag{14}$$

Note that

$$\xi_{2i} = (2\rho\kappa/\mu^2)(\cos\theta' - i\mu\cos\phi), \tag{15}$$

$$\xi_{li} = (2\rho\kappa/\mu^2)[(\epsilon_l/\epsilon_2) - (\sin\theta' - i\mu\sin\phi)^2]^{1/2} \qquad l = 1, 3. \tag{16}$$

In regions 1 and 3, ξ_{li} is either positive real (propagating wave) or negative imaginary (evanescent wave). The system of linear equations given by Eqs. (13) and (14) can be solved for C_m, and then R_i and T_i can be calculated from Eqs. (9) and (11). Note that the number of equations available is exactly equal to the number of unknowns. For example, if n waves are retained in the analysis, then there will be n unknown values each of R_i and of T_i and $2n$ unknown values of C_m. This is because the coefficient matrix $[b]$ in Eq. (7) is a $2n \times 2n$ matrix and therefore has $2n$ eigenvalues, and thus there are $2n$ unknown values of C_m. Alternatively, this may be viewed as being due to the n coupled-wave equations, each being a second-order differential equation, and thus there are $2n$ roots or eigenvalues and $2n$ unknown constants C_m to be determined from the boundary conditions. Therefore the total number of unknowns is $4n$, and Eqs. (9)–(12) provide $4n$ linear equations in these unknowns.

To summarize, the algorithm used to solve this problem proceeds as follows: First the coefficient matrix $[b]$ is constructed, and then eigenvalues and eigenvectors are calculated (typically by using a computer library program). The system of linear equations, Eqs. (13) and (14), is then constructed and solved for C_m (using a technique such as gauss elimination). Equations (9) and (11) are then used to calculate the diffracted amplitudes R_i and T_i. Power conservation requires that the sum of the efficiencies for all of the propagating waves be unity. That is,

$$\sum_i (DE_{1i} + DE_{3i}) = 1, \tag{17}$$

where DE_{1i} and DE_{3i} are the diffraction efficiencies in regions 1 and 3, respectively. These diffraction efficiencies are given by

$$DE_{1i} = Re(\xi_{1i}/\xi_{10})R_iR_i^* \tag{18}$$

and

$$DE_{3i} = Re(\xi_{3i}/\xi_{10})T_iT_i^*. \tag{19}$$

The real part of the ratio of the propagation constants occurs when the time-average power-flow density is obtained by taking the real part of the complex Poynting vector. The quantity $\mathrm{Re}(\xi_{li}/\xi_{10})$ is just the usual ratio of the cosine of the diffraction angle for the ith wave to the cosine of the angle of the zeroth-order wave for the lth medium. The results of sample calculations for the diffraction efficiencies are shown in Figs. 2–6 as a function of the grating strength parameter γ ($= \kappa d/\cos\theta'$) and as a function of d/Λ for a pure transmission grating ($\phi = \pi/2$), for general slanted gratings ($\phi = \pi/3, \pi/6$), and for a pure reflection grating ($\phi = 0$).

4. COMPARISON WITH PREVIOUS ANALYSES

A. Rigorous Modal Approach

Basically, the rigorous coupled-wave approach presented in this paper and the rigorous modal approach analyze the pla-

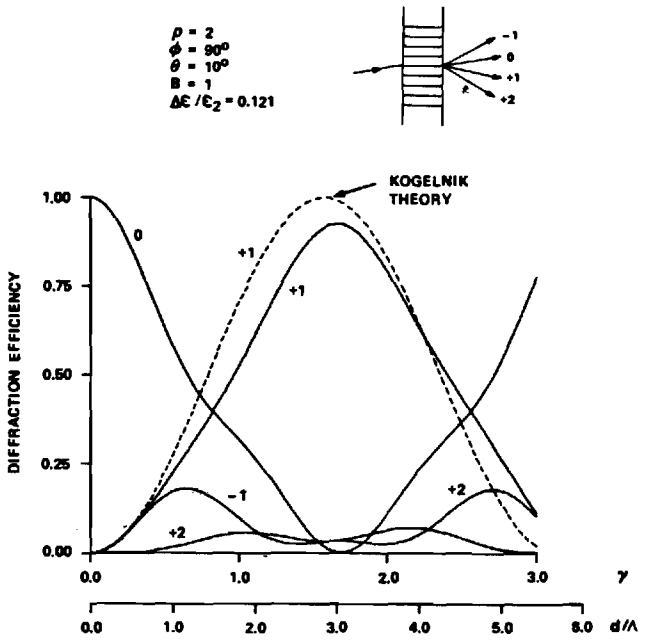


Fig. 2. The diffraction efficiencies of the transmitted waves for pure transmission grating ($\phi = 90°$), with $\rho = 2$. The diffraction efficiencies of all reflected and transmitted waves not shown in the figure are less than 0.01. Here and in Figs. 3–6, $\epsilon_1 = \epsilon_2 = \epsilon_3$.

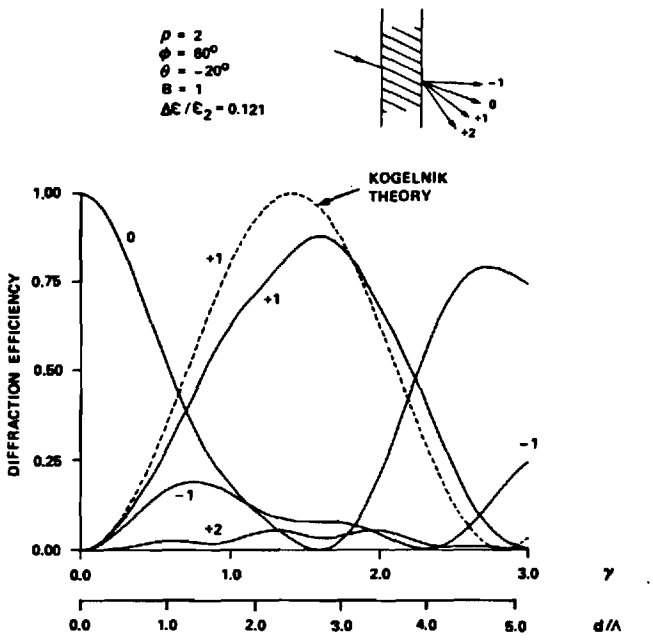


Fig. 3. The diffraction efficiencies of the transmitted waves for a $\phi = 60°$ slanted grating with $\rho = 2$. The diffraction efficiencies of all reflected and transmitted waves not shown in the figure are less than 0.01.

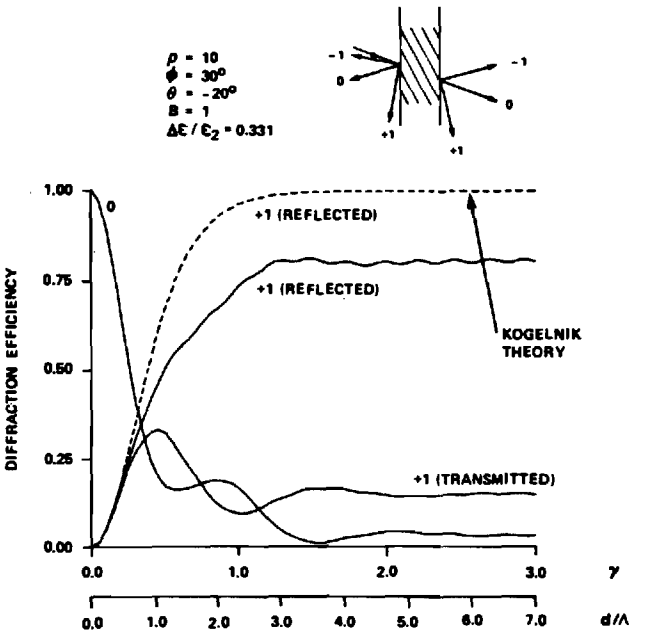


Fig. 4. The diffraction efficiencies for a $\phi = 30°$ grating with $\rho = 10$. The diffraction efficiencies of all reflected and transmitted waves not shown in the figure are less than 0.01.

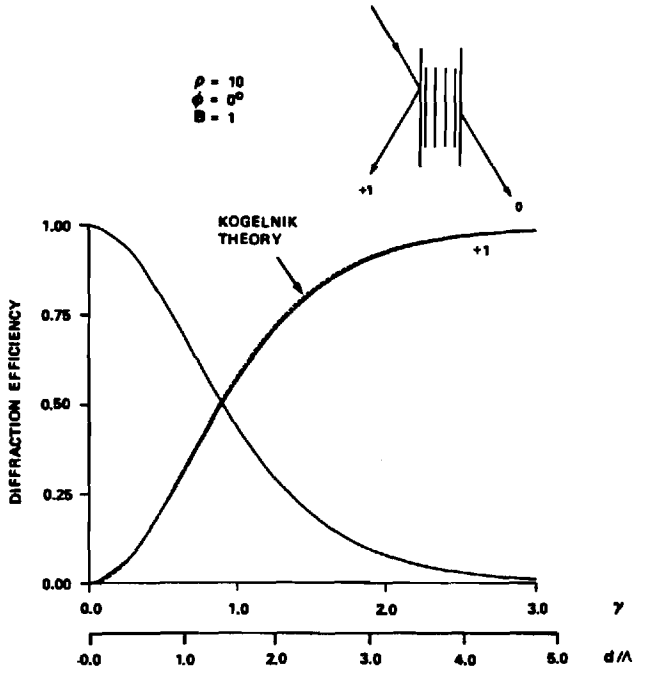


Fig. 5. The diffraction efficiencies of the transmitted and reflected waves at first-Bragg incidence ($B = 1$) for a pure reflection grating ($\phi = 0°$) with $\rho = 10$. These two waves are the only waves that propagate in the unmodulated regions.

nar-grating diffraction problem by solving the wave equation in the three regions (Fig. 1) and then matching the tangential electric and the magnetic fields at the two boundaries to determine all the unknowns. The main difference between the two approaches is in the technique used to find solutions of the wave equation in the modulated region. In the present coupled-wave approach, the resulting system of coupled-wave equations is formulated into a simple matrix form for which the solution is readily obtained by calculating the eigenvalues and the eigenvectors of the coefficient matrix constructed from the coupled-wave equations. The eigenvalue problem, although not simple, has been extensively studied, and numerous efficient and straightforward computer programs to calculate the eigenvalues and the eigenvectors are available in typical computer-program libraries. The modal approach, on the other hand, requires that a transcendental relationship

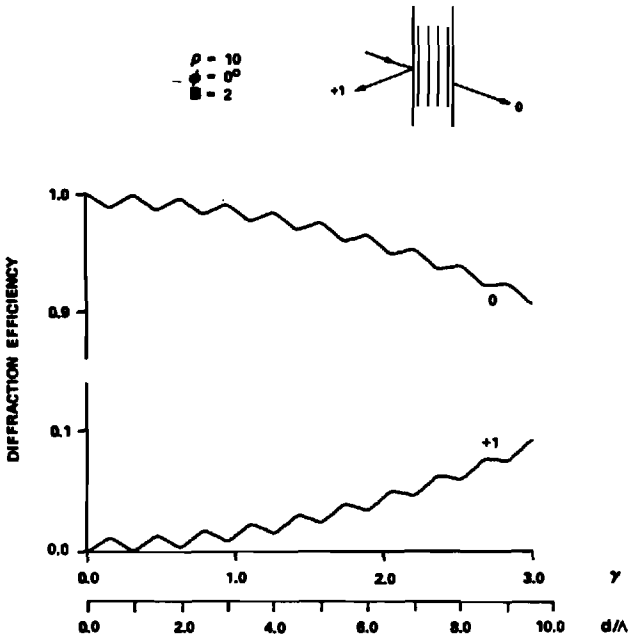


Fig. 6. The diffraction efficiencies of the transmitted and reflected waves at second-Bragg incidence ($B = 2$) for a pure reflection grating ($\phi = 0°$) with $\rho = 10$. These two waves are the only waves that propagate in the unmodulated regions.

in the form of a continued fraction expansion be solved to find the wavenumbers and their corresponding coefficients that are needed to solve the wave equation in the modulated region. This primary difficulty when applying the modal approach to the analysis of slanted gratings is due to the fact that there is no systematic technique yet available to solve this transcendental continued fraction relation for the general slanted-grating case.[8] For the modal approach the transcendental relationship can also be formulated as a matrix. This matrix is always $n \times n$ (as opposed to $2n \times 2n$ for the present coupled-wave analysis). Only for the unslanted, physically symmetric case is the resulting matrix for the modal approach in standard eigenvalue form. The corresponding wave numbers needed in the solution are the positive and negative square roots of the eigenvalues. The vector of coefficients, which is the same for both the positive and negative wave numbers, is the corresponding eigenvector of the problem. For slanted gratings the resulting modal-approach $n \times n$ matrix is not in the form of a standard eigenvalue problem. Therefore the wave numbers occurring in the matrix cannot be systematically determined.

B. Approximate Modal and Coupled-Wave Analyses

With the exception of the rigorous modal theory, previous analyses have typically been approximate. The present coupled-wave analysis is exact and rigorous, and it reduces to each of the previous analyses with the appropriate approximations and simplifications. For example, it reduces to Kong's coupled-wave analysis[6] and to Bergstein and Kermisch's analysis,[12] if a two-wave regime is assumed (that is, retaining only the first- and zero-order waves and neglecting all higher-order waves). Kogelnik's[4] analysis is obtained by assuming a two-wave regime, neglecting boundary diffraction, and neglecting second derivatives of the field amplitudes in Eq. (6). This last approximation, which is common in almost all previous coupled-wave analyses,[1-5,7] implies that the parameter $\Delta\epsilon/8\epsilon_2$ was assumed to be very small in these analyses. Magnusson and Gaylord's[7] analysis is obtained by neglecting boundary diffraction and the second derivatives of the field amplitudes. It is important to note that in previous coupled-wave analyses,[1-5,7] the amplitudes and intensities of the diffracted waves were calculated inside the modulated region. These analyses are based on solving the half-space grating problem. They do not (and cannot, because of the various approximations) solve the problem of general planar slab grating bounded by two different media.

C. Evaluation of Approximate Approaches

The rigorous coupled-wave theory presented in this paper is easily implemented and permits calculations to be performed rapidly. As such, this analysis is a powerful tool for analyzing the precise effect of using the various assumptions employed in the above approximate methods. It is difficult to quantify the error in general terms. However, the *trends* shown here have been consistently observed in the results of numerous calculations. Figures 7–10 show, for some typical cases, the absolute error in the primary +1 order that is due to using (a) the two-wave coupled-wave theory,[4] (b) the multiwave coupled-wave theory,[7] and (c) the two-wave modal theory.[12] In general, these grating structures do not behave simply as transmission gratings or reflection gratings. This is most obvious in slanted gratings (such as that shown in Fig. 4). A combination of transmission and reflection behaviors is typically present, even though one or the other dominates in each case, as determined by the location of the primary +1 diffracted order. For the slanted grating of Fig. 4, the +1 reflected order is the primary +1 diffracted order, and reflection behavior dominates in this case. Note that the +1 transmitted wave (which is phase matched to the +1 reflected wave) would have been assumed to be zero in the two-wave and multiwave (approximate) coupled-wave theories. This nonzero +1 transmitted diffracted wave is attributable entirely to the inclusion of second derivatives in this rigorous theory.

The gratings represented in Figs. 7–10 include the parameter values used in Figs. 2–5. The effect of neglecting boundary diffraction and second derivatives of field amplitudes is determined by the absolute error that results from using multiwave coupled-wave theory. As is shown in Figs. 7 and 8, this error is relatively small when forward-diffracted waves are dominant (transmission-grating behavior). However, when backward-diffracted waves are dominant (reflection-grating behavior), this error becomes relatively large, as is shown in Fig. 9. The effect of neglecting higher-order waves and thus retaining only one diffracted wave is determined by the absolute error that results in using two-wave modal theory. In a complementary way to the preceding results, Figs. 7 and 8 show that this error is large when forward-diffracted waves are dominant, and Fig. 9 shows that it is small when backward-diffracted waves are dominant. The results of many calculations support these conclusions. For example, calculations for two slanted gratings with exactly the same modulation, but one with transmission behavior dominant and the other with reflection behavior dominant, show that higher-order waves are more important in the transmission case and second derivatives and boundary diffraction are more important in the reflection case. Further, these basic conclu-

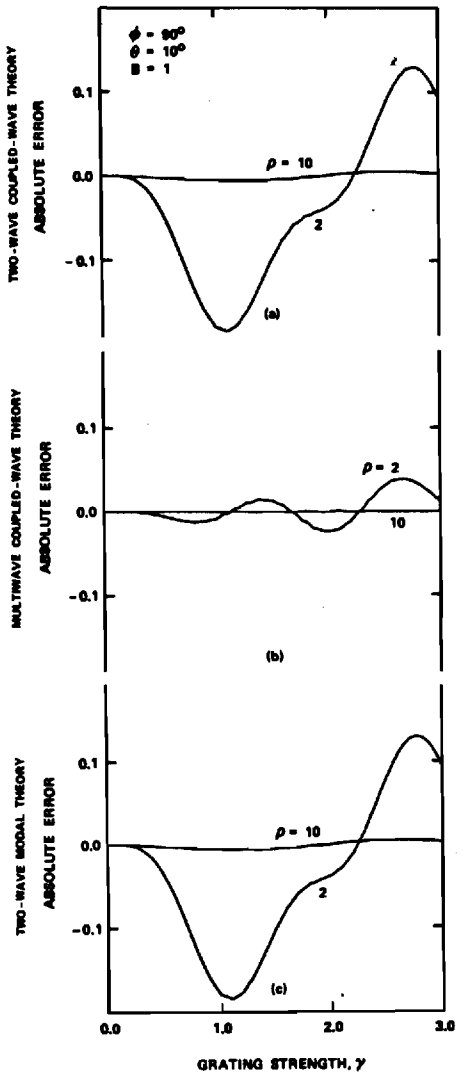


Fig. 7. Absolute error in the first-order diffraction efficiency as predicted by (a) two-wave coupled-wave theory, (b) multiwave coupled-wave theory, and (c) two-wave modal theory for the grating conditions shown in Fig. 2.

sions are supported by calculations for slanted gratings of equal modulation over a wide range of grating modulations. For example, at very small modulations, even though the errors that are due to the neglect of higher-order waves are small and the errors that are due to the neglect of second derivatives and boundary diffraction are small, the importance of these assumptions relative to each other is exactly the same as for large modulations. For equal grating modulations, higher-order waves are more important when transmission behavior dominates, and second derivatives and boundary diffraction are more important when reflection behavior dominates. This relative importance is unchanged regardless of whether the calculated errors are large or small. The combined effect of neglecting both higher-order waves and second derivatives and boundary diffraction is determined by the absolute error that results from using two-wave coupled-wave theory. This error is also shown in Figs. 7–9. From the two-wave coupled-wave theory error, it is consistently found that the total error that is due to neglecting both higher-order waves and second derivatives and boundary diffraction is approximately equal to the larger of the two constituent errors. The errors shown in Figs. 7–9 are all for the primary +1 order. However, numerous additional calculations for the other waves show a similar dependence on the approximations. For a large variety of cases analyzed, it was found in every case that if higher-order waves or boundary diffraction and second derivatives are important in accurately calculating the primary +1 order, then they are similarly important in accurately calculating the other orders.

For the case of a pure reflection grating ($\phi = 0$), the situation is somewhat more complicated. As ϕ approaches zero, all the forward-diffracted and all the backward-diffracted waves outside the modulated region converge to a single forward and a single backward direction, respectively. The directions of the various orders inside the modulated medium remain distinct. The error associated with the total backward-diffracted wave is shown in Fig. 10. In this case, the errors that are due to neglect of second derivatives and boundary diffraction are similar to those that are due to neglect of higher-order waves.

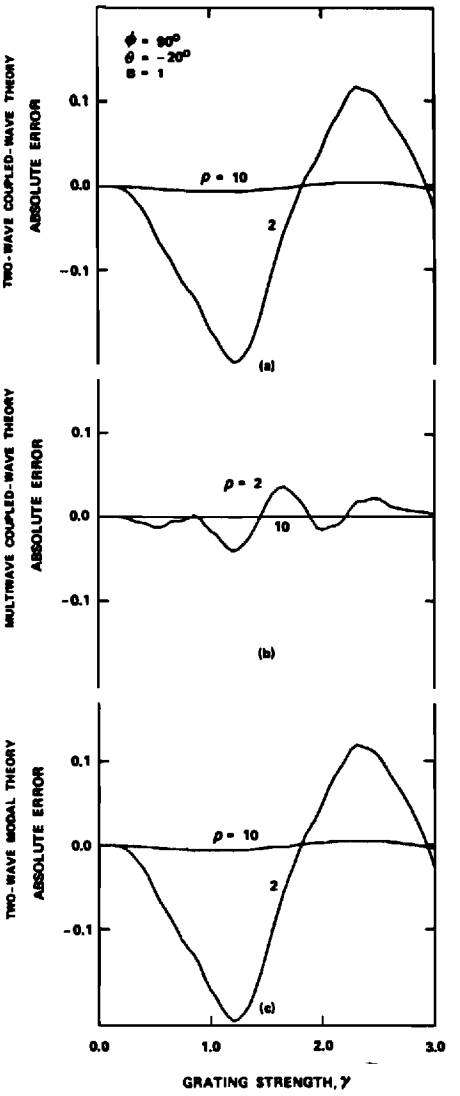


Fig. 8. Absolute error in the first-order diffraction efficiency as predicted by (a) two-wave coupled-wave theory, (b) multiwave coupled-wave theory, and (c) two-wave modal theory for the grating conditions shown in Fig. 3.

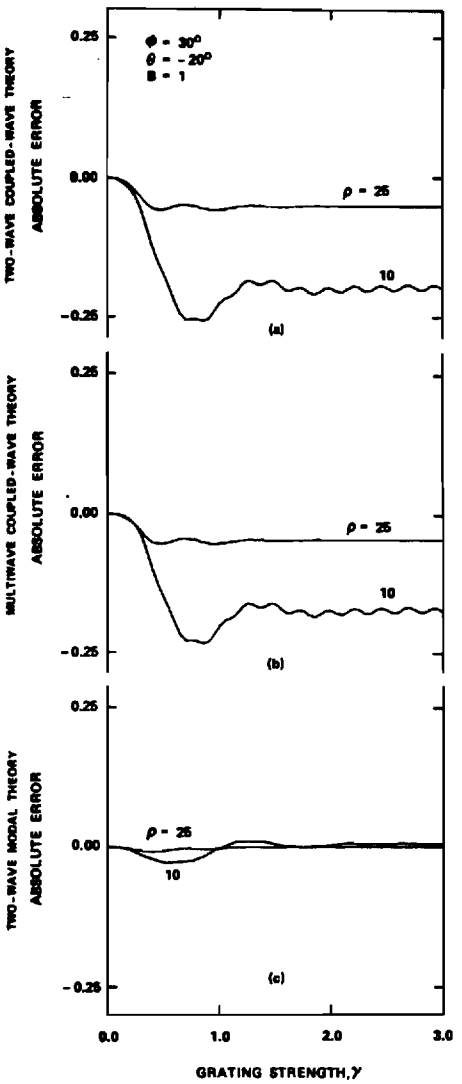


Fig. 9. Absolute error in the first-order diffraction efficiency as predicted by (a) two-wave coupled-wave theory, (b) multiwave coupled-wave theory, and (c) two-wave modal theory for the grating conditions shown in Fig. 4.

Incidence at the second-order Bragg condition is shown for a pure reflection grating in Fig. 6. It is important to note that solutions are not possible at even-order Bragg incidence ($B = 2, 4, 6, \ldots$) unless second derivatives and boundary diffraction are included in the analysis. For even-order Bragg conditions the coefficient of the first derivative term in Eq. (6) vanishes for pure reflection gratings. Thus calculations such as those shown in Fig. 6 are not even possible with most approximate theories.

5. DISCUSSION AND CONCLUSIONS

The diffraction of a plane electromagnetic wave incident obliquely upon a planar grating bounded by two different media has been analyzed by using an exact and rigorous coupled-wave approach. The grating vector may have any arbitrary angle with respect to the boundaries. The solution has been formulated in terms of state variables in a simple matrix form that is easily implemented on a digital computer. Sample rigorous calculations have been presented for transmission, reflection, and, for the first time, for general slanted gratings.

The present coupled-wave approach has been compared with the rigorous modal approach and with approximate modal and coupled-wave approaches. It was shown that inclusion of higher-order waves is more important than the inclusion of second derivatives and boundary diffraction for obtaining accurate predictions of diffraction efficiency when forward-diffracted waves are dominant (transmission-grating behavior). Conversely, it was shown that inclusion of second derivatives of the field amplitudes and boundary diffraction is more important for obtaining accurate predictions of diffraction efficiency when backward-diffracted waves are dominant (reflection-grating behavior).

The present method of analysis is useful in all applications in which planar gratings are utilized. However, it is especially valuable in applications in which slanted gratings and reflection gratings are used, such as in grating couplers and distributed-feedback lasers, since it is difficult to use the modal approach in these cases. This analysis may also be

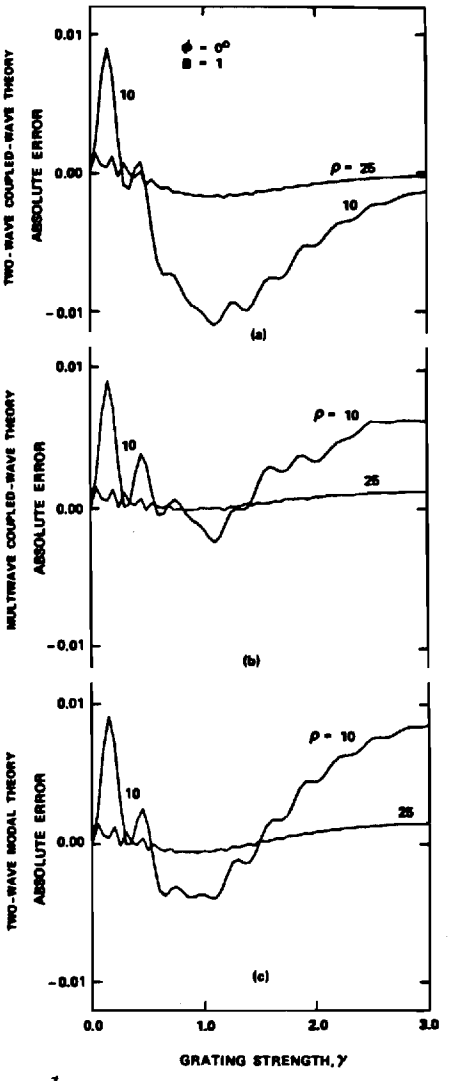


Fig. 10. Absolute error in the first-order diffraction efficiency as predicted by (a) two-wave coupled-wave theory, (b) multiwave coupled-wave theory, and (c) two-wave modal theory for the grating conditions shown in Fig. 5.

straightforwardly extended to absorption gratings and to mixed phase and absorption gratings, if desired.

APPENDIX: THE STATE EQUATION

The elements of the matrix $[b_{rs}]$ in Eq. (7), the state equation, are determined by Eq. (6), the governing set of differential equations. By defining the state variables as

$$S_{1,i} = S_i(u), \tag{A1}$$

$$S_{2,i} = \dot{S}_i(u), \tag{A2}$$

the constituent state equations may be written as

$$\dot{S}_{1,i} = S_{2,i}, \tag{A3}$$

$$\dot{S}_{2,i} = aS_{1,i+1} + b_iS_{1,i} + aS_{1,i-1} + c_iS_{2,i}, \tag{A4}$$

where $a = 8\epsilon_2/\Delta\epsilon$, $b_i = -a\rho i(i - B)$, and $c_i = a(\cos\theta' - i\mu\cos\phi)$. Writing Eqs. (A3) and (A4) as a single matrix state equation gives Eq. (7). In expanded form it is

$$\begin{bmatrix} \vdots \\ \dot{S}_{1,2} \\ \dot{S}_{1,1} \\ \dot{S}_{1,0} \\ \dot{S}_{1,-1} \\ \dot{S}_{1,-2} \\ \vdots \\ \dot{S}_{2,2} \\ \dot{S}_{2,1} \\ \dot{S}_{2,0} \\ \dot{S}_{2,-1} \\ \dot{S}_{2,-2} \\ \vdots \end{bmatrix} = \begin{bmatrix} & & \vdots & & & & & \vdots & & \\ & 0 & 0 & 0 & 0 & & 1 & 0 & 0 & 0 & 0 & \\ & 0 & 0 & 0 & 0 & & 0 & 1 & 0 & 0 & 0 & \\ \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 & 0 & \cdots \\ & 0 & 0 & 0 & 0 & & 0 & 0 & 0 & 1 & 0 & \\ & 0 & 0 & 0 & 0 & & 0 & 0 & 0 & 0 & 1 & \\ & & & \vdots & & & & & \vdots & & & \\ & b_2 & a & 0 & 0 & & c_2 & 0 & 0 & 0 & 0 & \\ & a & b_1 & a & 0 & & 0 & c_1 & 0 & 0 & 0 & \\ \cdots & 0 & a & b_0 & a & \cdots & 0 & 0 & c_0 & 0 & 0 & \cdots \\ & 0 & 0 & a & b_{-1} & & 0 & 0 & 0 & c_{-1} & 0 & \\ & 0 & 0 & 0 & a & b_{-2} & 0 & 0 & 0 & 0 & c_{-2} & \\ & & & \vdots & & & & & \vdots & & & \end{bmatrix} \begin{bmatrix} \vdots \\ S_{1,2} \\ S_{1,1} \\ S_{1,0} \\ S_{1,-1} \\ S_{1,-2} \\ \vdots \\ S_{2,2} \\ S_{2,1} \\ S_{2,0} \\ S_{2,-1} \\ S_{2,-2} \\ \vdots \end{bmatrix}. \tag{A5}$$

This work was sponsored by the National Science Foundation under grant no. ECS-7919592 and by the Joint Services Electronics Program under grant no. DAAG29-78-C-0005.

REFERENCES

1. R. R. Aggrawal, "Diffraction of light by ultrasonic waves," Proc. Indian Acad. Sci. **31**, 417–426 (1950).
2. P. Phariseau, "On the diffraction of light by progressive supersonic waves," Proc. Indian Acad. Sci. Sect. A **44**, 165–170 (1965).
3. W. R. Klein and B. D. Cook, "Unified approach to ultrasonic light diffraction," IEEE Trans. Sonic Ultrason. **SU-14**, 123–134 (1967).
4. H. Kogelnik, "Coupled wave theory for thick hologram gratings," Bell Syst. Tech. J. **48**, 2909–2947 (1969).
5. G. L. Fillmore and R. F. Tynan, "Sensitometric characteristics of hardened dichromated-gelatin films," J. Opt. Soc. Am. **61**, 199–203 (1971).
6. J. A. Kong, "Second-order coupled-mode equations for spatially periodic media," J. Opt. Soc. Am. **67**, 825–829 (1977).
7. R. Magnusson and T. K. Gaylord, "Analysis of multiwave diffraction by thick gratings," J. Opt. Soc. Am. **67**, 1165–1170 (1977).
8. T. Tamir, H. C. Wang, and A. A. Oliner, "Wave propagation in sinusoidally stratified dielectric media," IEEE Trans. Microwave Theory Tech. **MTT-12**, 323–335 (1964).
9. T. Tamir and H. C. Wang, "Scattering of electromagnetic waves by a sinusoidally stratified half-space: I. Formal solution and analysis approximations," Can. J. Phys. **44**, 2073–2094 (1966).
10. T. Tamir, "Scattering of electromagnetic waves by a sinusoidally stratified half-space: II. Diffraction aspects of the Rayleigh and Bragg wavelengths," Can. J. Phys. **44**, 2461–2494 (1966).
11. C. B. Burckhardt, "Diffraction of a plane wave at a sinusoidally stratified dielectric grating," J. Opt. Soc. Am. **56**, 1502–1509 (1966).
12. L. Bergstein and D. Kermisch, "Image storage and reconstruction in volume holography," Proc. Symp. Modern Opt. **17**, 655–680 (1967).
13. R. S. Chu and T. Tamir, "Guided-wave theory of light diffraction by acoustic microwaves," IEEE Trans. Microwave Theory Tech. **MTT-18**, 486–504 (1970).
14. R. S. Chu and T. Tamir, "Wave propagation and dispersion in space–time periodic media," Proc. IEE **119**, 797–806 (1972).
15. F. G. Kaspar, "Diffraction by thick periodically stratified gratings with complex dielectric constant," J. Opt. Soc. Am. **63**, 37–45 (1973).
16. S. T. Peng, T. Tamir, and H. L. Bertoni, "Theory of periodic dielectric waveguides," IEEE Trans. Microwave Theory Tech. **MTT-23**, 123–133 (1975).
17. R. S. Chu and J. A. Kong, "Modal theory of spatially periodic media," IEEE Trans. Microwave Theory Tech. **MTT-25**, 18–24 (1977).
18. R. Magnusson and T. K. Gaylord, "Equivalence of multiwave coupled-wave theory and modal theory of periodic-media diffraction," J. Opt. Soc. Am. **68**, 1777–1779 (1978).
19. M. G. Moharam and T. K. Gaylord, "Coupled-wave analysis of reflection gratings," Appl. Opt., **20**, 240–244 (1981).
20. M. G. Moharam, T. K. Gaylord, and R. Magnusson, "Criteria for Bragg regime diffraction by phase gratings," Opt. Commun. **32**, 14–18 (1980).

Rigorous coupled-wave analysis of grating diffraction—*E*-mode polarization and losses

M. G. Moharam and T. K. Gaylord

School of Electrical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

Received August 21, 1982

Rigorous coupled-wave theory of diffraction by dielectric gratings is extended to cover *E*-mode polarization and losses. Unlike in the *H*-mode-polarization case, it is shown that, in the *E*-mode case, direct coupling exists between all diffracted orders rather than just between adjacent orders.

INTRODUCTION

Optical diffraction by dielectric gratings has been the subject of extensive, sustained research for many years. Fields of application include acousto-optics, integrated optics, quantum electronics, holography, and spectroscopy. Grating-device functions include laser-beam deflection, modulation, coupling, filtering, distributed feedback, distributed Bragg reflection, holographic beam combining, wavelength multiplexing, wavelength demultiplexing, and others.

A rigorous coupled-wave theory (without approximations) has recently been formulated for dielectric gratings.[1] This analysis applies for incident light of *H*-mode polarization (electric field perpendicular to the plane of incidence and perpendicular to the grating vector). It is the purpose of this paper (1) to show how the rigorous coupled-wave analysis can be extended to treat *E*-mode polarization (electric field in the plane of incidence and in the plane of the grating vector) and lossy gratings, (2) to show that coupling exists between all diffracted orders for *E*-mode polarization (unlike the case for *H*-mode polarization in which the coupling is only between adjacent orders), and (3) to compare rigorous *E*-mode results for gratings with and without losses with previous approximate *E*-mode results, rigorous *H*-mode results, and approximate *H*-mode results.

GRATING WAVE EQUATIONS

General Vector Wave Equations

The lossy dielectric grating is characterized by a relative permittivity that is periodic and is given by

$$\epsilon(x, z) = \hat{\epsilon}_0 + \epsilon_1 \cos[K(x \sin \phi + z \cos \phi)], \tag{1}$$

where $\hat{\epsilon}_0$ is the average complex relative permittivity given by

$$\hat{\epsilon}_0 = \epsilon_0 - j\sigma_0/\omega\varepsilon_0, \tag{2}$$

ϵ_0 is the average relative permittivity, σ_0 is the average conductivity (representing the nonspatially varying losses), ω is the optical radian frequency, ε_0 is the permittivity of free space, ϵ_1 is the amplitude of the sinusoidal relative permittivity, ϕ is the grating slant angle, K is the magnitude of the grating vector given by $K = 2\pi/\Lambda$, and Λ is the grating period. The planar boundaries of the grating are perpendicular to the z direction at $z = 0$ and $z = d$. Although Eq. (1) represents the particular case of a sinusoidal permittivity, other grating profiles can also be treated. The electromagnetic fields inside a planar lossy dielectric grating with a spatially varying relative permittivity are given by vector wave equations obtained directly from Maxwell's equations. The electric-field vector wave equation is

$$\nabla^2\overline{E} + \nabla\left(\overline{E} \cdot \frac{\nabla\epsilon}{\epsilon}\right) + k^2\epsilon(x, z)\overline{E} = 0, \tag{3}$$

where $\overline{E}$ is the electric field, $\epsilon(x, z)$ is the periodic complex relative permittivity (dielectric constant) $k = 2\pi/\lambda$, and λ is the free-space wavelength. Similarly, the magnetic-field vector wave equation is

$$\nabla^2\overline{H} + \frac{\nabla\epsilon}{\epsilon} \times \nabla \times \overline{H} + k^2\epsilon(x, z)\overline{H} = 0, \tag{4}$$

where $\overline{H}$ is the magnetic field. These general wave equations may be considerably simplified for particular incident wave polarizations.

H-Mode-Polarization Wave Equation

For *H*-mode polarization (electric field perpendicular to plane of incidence and perpendicular to the grating vector), the electric field is solely in the y direction, and so $\overline{E} = E\hat{y}$, where $\hat{y}$ is the unit vector in the y direction. Because the electric field is perpendicular to the grating modulation vector, then $\overline{E} \cdot \nabla\epsilon = 0$. Electric-field vector wave Eq. (3) therefore reduces to the scalar Helmholtz wave equation

$$\nabla^2 E + k^2\epsilon(x, z)E = 0. \tag{5}$$

This is the equation that is commonly solved in the analysis of dielectric grating diffraction.

E-Mode-Polarization Wave Equation

For *E*-mode polarization, the electric field is in the plane of incidence, and this plane contains the grating vector. The magnetic field is solely in the y direction, and so $\overline{H} = H\hat{y}$. Because the magnetic field is only in the y direction, it is advantageous to select and to work with the magnetic-field vector wave Eq. (4). This vector wave equation may be simplified by using the vector identities $\nabla\epsilon \times \nabla \times \overline{H} \equiv \nabla(\nabla\epsilon \cdot \overline{H}) - (\nabla\epsilon \cdot \nabla)\overline{H} - (\overline{H} \cdot \nabla)\nabla\epsilon - \overline{H} \times (\nabla \times \nabla\epsilon)$ and $\nabla \times \nabla\epsilon \equiv 0$.

For E-mode polarization, H is perpendicular to $\nabla\epsilon$, and thus $\nabla\epsilon \cdot \overline{H} = 0$ and $(\overline{H} \cdot \nabla)\nabla\epsilon = 0$. The magnetic-field vector wave equation thus reduces to

$$\nabla^2 H - \left(\frac{\nabla\epsilon}{\epsilon} \cdot \nabla\right) H + k^2 \epsilon(x,z) H = 0. \quad (6)$$

This equation contains an additional term in comparison with E-mode wave Eq. (5).

COUPLED-WAVE EQUATIONS

H-Mode Coupled-Wave Equations

The H-mode polarization coupled-wave equations may be obtained by expanding the electric field in space harmonics as[1]

$$E(x,z) = \sum_{i=-\infty}^{+\infty} S_i(z) \exp(-j\overline{\sigma}_i \cdot \overline{r}), \quad (7)$$

where i is the integer space-harmonic index, $S_i(z)$ is the space-harmonic electric-field amplitude, $\overline{\sigma}_i = \overline{k}_2 - i\overline{K}$ from the Floquet theorem, and $\overline{k}_2$ is the wave vector of the zero-order ($i = 0$) refracted wave in region 2, the grating region $0 \le z \le d$. (Region 1 is the input region $z \le 0$, and region 3 is the output region $z \ge d$.) The magnitude of $\overline{k}_2$ is $k_2 = 2\pi(\hat{\epsilon}_0)^{1/2}/\lambda$. Each space harmonic $S_i(z)$ inside the grating is phase matched to a forward-diffracted and a backward-diffracted wave. These waves may be either propagating or evanescent. Substitution of Eq. (7) into Eq. (5) leads to an infinite exponential series in terms of $S_i(z)$. Each coefficient may be expressed as a function of i and z, and each exponent as a function of i and x. For nontrivial solutions, each coefficient must be equal to zero. This gives the coupled-wave equations. After simplification, the H-mode coupled-wave equations are

$$\frac{1}{2\pi^2}\frac{d^2 S_i(z)}{dz^2} - j\frac{2}{\pi}\left[\frac{(\hat{\epsilon}_0 - \epsilon_I \sin\theta')^{1/2}}{\lambda} - \frac{i\cos\phi}{\Lambda}\right]\frac{dS_i(z)}{dz} + \frac{2i(m-i)}{\Lambda^2} S_i(z) + \frac{\epsilon_1}{\lambda^2}[S_{i+1}(z) + S_{i-1}(z)] = 0, \quad (8)$$

where θ' is the angle of incidence in region 1 of the input plane wave, ϵ_I is the average relative permittivity in region 1, and m is defined as

$$m \equiv 2(\Lambda/\lambda)[\epsilon_I{}^{1/2} \sin\phi \sin\theta' + (\hat{\epsilon}_0 - \epsilon_I \sin^2\theta')^{1/2} \cos\phi]. \quad (9)$$

When the real part of m is an integer, this represents a Bragg condition. These rigorous coupled-wave equations may be solved by the state variable methods,[2] and, together with the appropriate boundary conditions, all the diffracted fields may be determined.[1]

E-Mode Coupled-Wave Equations

The vectorial E-mode wave Eq. (6) can also be reformulated as a set of scalar coupled-wave equations. The vector term may be expanded as

$$-\left(\frac{\nabla\epsilon}{\epsilon} \cdot \nabla\right) H = \frac{\epsilon_1 \sin(\overline{K} \cdot \overline{r})}{\epsilon_0 + \epsilon_1 \cos(\overline{K} \cdot \overline{r})}\left(\sin\phi \frac{\partial H}{\partial x} + \cos\phi \frac{\partial H}{\partial z}\right)\frac{2\pi}{\Lambda}, \quad (10)$$

and thus only a y-component equation exists. To put this term into standard form, the leading factor is expanded in a Fourier series as

$$\frac{\epsilon_1 \sin(\overline{K} \cdot \overline{r})}{\epsilon_0 + \epsilon_1 \cos(\overline{K} \cdot \overline{r})} = -j \sum_{h=-\infty}^{+\infty} A_h \exp(jh\overline{K} \cdot \overline{r}), \quad (11)$$

where $A_h = -\{[(\hat{\epsilon}_0/\epsilon_1)^2 - 1]^{1/2} - (\hat{\epsilon}_0/\epsilon_1)\}^h$ for $h \ge 1$, $A_{-h} = -A_h$, and $A_0 = 0$. For the E-mode case, the magnetic field is expanded in space harmonics as

$$H(x,z) = \sum_{i=-\infty}^{+\infty} U_i(z) \exp(-j\overline{\sigma}_i \cdot \overline{r}), \quad (12)$$

where $U_i(z)$ is the space-harmonic magnetic-field amplitude and the other quantities are defined as before. Substituting Eqs. (10)–(12) into Eq. (6) and proceeding as before gives the E-mode coupled-wave equations as

$$\frac{1}{2\pi^2}\frac{d^2 U_i(z)}{dz^2} - j\frac{2}{\pi}\left[\frac{(\hat{\epsilon}_0 - \epsilon_I \sin\theta')^{1/2}}{\lambda} - \frac{i\cos\phi}{\Lambda}\right]\frac{dU_i(z)}{dz} - j\frac{\cos\phi}{\pi\Lambda}\sum_h A_h \frac{dU_{i-h}(z)}{dz} + \frac{2i(m-i)}{\Lambda^2} U_i(z) + \frac{\epsilon_1}{\lambda^2}[U_{i+1}(z) + U_{i-1}(z)] + \frac{2}{\Lambda^2}\sum_h \left(i - h - \frac{m}{2}\right) \times A_h U_{i-h}(z) = 0. \quad (13)$$

These equations for E-mode polarization are clearly more complicated than H-mode coupled-wave Eqs. (8). The two additional terms in Eq. (13) both contain a series in A_h and U_{i-h}. Whereas the H-mode equations contain only S_{i-1}, S_i, and S_{i+1} amplitude terms, the E-mode equations contain U_{i-1}, U_i, U_{i+1}, and U_{i-h} amplitude terms. Therefore in the H-mode case there is direct coupling only between adjacent orders, but in the E-mode case there is direct coupling among all diffracted orders. Although the number of terms in the E-mode case is larger, the resulting coupled-wave equations may be solved by the state-variables method in exactly the same manner as in the H-mode case.

BOUNDARY CONDITIONS

At the boundaries of the grating ($z = 0$ and $z = d$), the tangential components of the electric field and the magnetic field must be continuous. In this way, the field of each diffracted order outside the grating volume is related to the corresponding space-harmonic field inside the grating. Thus, in order to construct the boundary conditions, the tangential components of E and H must be determined.

H-Mode-Polarization Tangential Fields

For H-mode polarization, the tangential component of the electric field is the y component of E, and it is given by Eq. (7) directly. The values of $S_i(0)$ and $S_i(d)$ needed in Eq. (7) are obtained by solving H-mode coupled-wave Eq. (8) for $S_i(z)$. The tangential component of the magnetic field is the x component of H. It may be obtained from the Maxwell curl equation $\nabla \times \overline{E} = -\partial\overline{B}/\partial t$. The result is $H_x = (-j/\omega\mu)\partial E_y/\partial z$, and, together with Eq. (7), the tangential magnetic field is

$$H_x = (-j/\omega\mu)\frac{\partial}{\partial z}\sum_{i=-\infty}^{+\infty} S_i(z) \exp(-j\overline{\sigma}_i \cdot \overline{r}). \quad (14)$$

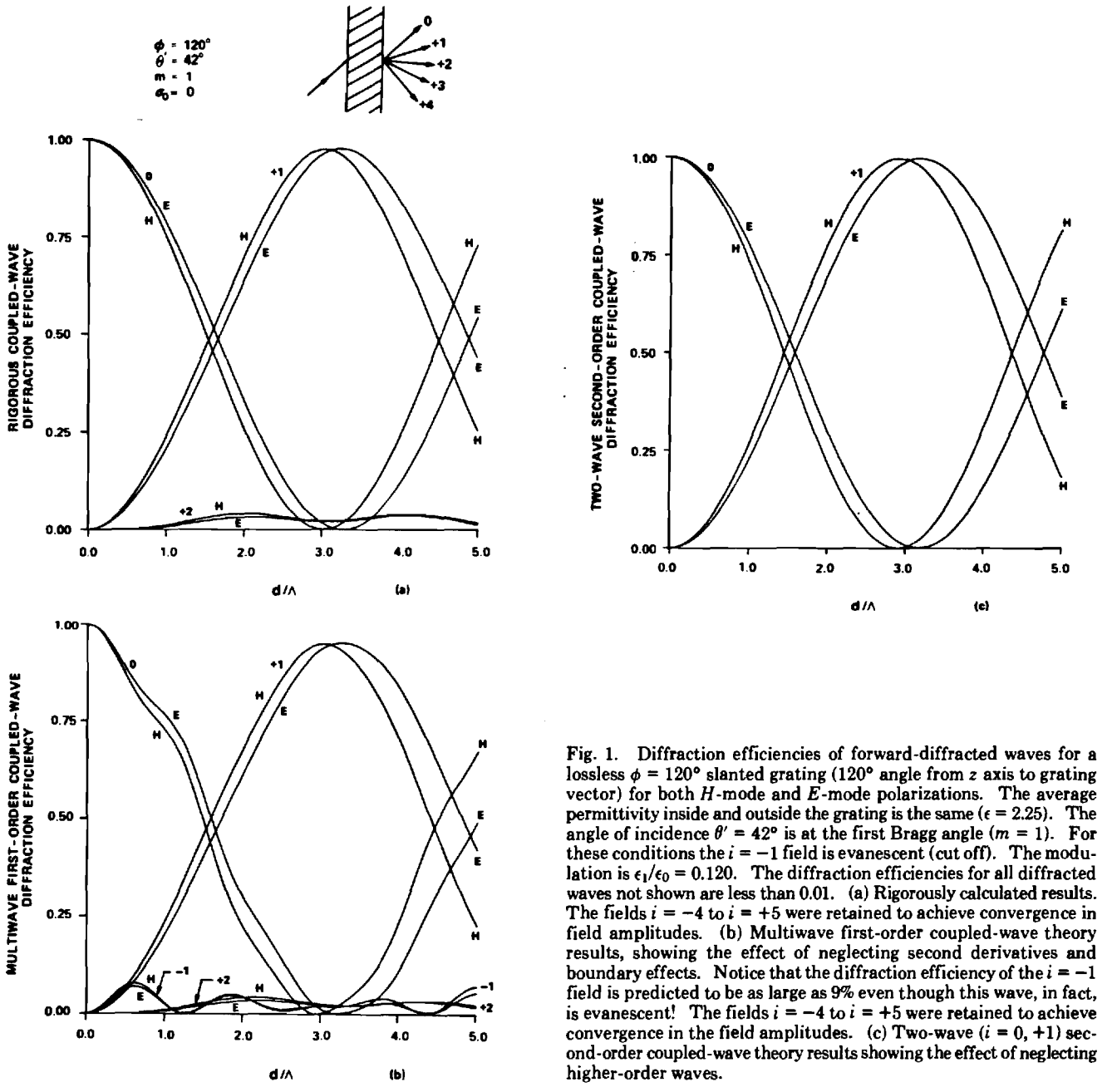


Fig. 1. Diffraction efficiencies of forward-diffracted waves for a lossless $\phi = 120°$ slanted grating (120° angle from z axis to grating vector) for both H-mode and E-mode polarizations. The average permittivity inside and outside the grating is the same ($\epsilon = 2.25$). The angle of incidence $\theta' = 42°$ is at the first Bragg angle ($m = 1$). For these conditions the $i = -1$ field is evanescent (cut off). The modulation is $\epsilon_1/\epsilon_0 = 0.120$. The diffraction efficiencies for all diffracted waves not shown are less than 0.01. (a) Rigorously calculated results. The fields $i = -4$ to $i = +5$ were retained to achieve convergence in field amplitudes. (b) Multiwave first-order coupled-wave theory results, showing the effect of neglecting second derivatives and boundary effects. Notice that the diffraction efficiency of the $i = -1$ field is predicted to be as large as 9% even though this wave, in fact, is evanescent! The fields $i = -4$ to $i = +5$ were retained to achieve convergence in the field amplitudes. (c) Two-wave ($i = 0, +1$) second-order coupled-wave theory results showing the effect of neglecting higher-order waves.

E-Mode-Polarization Tangential Fields

For E mode polarization, the tangential component of the magnetic field is the y component of H, and it is given by Eq. (12) directly. The values of $U_i(0)$ and $U_i(d)$ to be used in Eq. (12) are obtained by solving E-mode coupled-wave Eq. (13) for $U_i(z)$. The tangential electric field is the x component of E. It may be obtained from the other Maxwell curl equation $\nabla \times \bar{H} = \partial \bar{D}/\partial t$. The result is $E_x = [j/\omega\epsilon_0\epsilon(x, z)]\partial H_y/\partial z$. Expanding $1/\epsilon(x, z)$ into a Fourier series gives

$$\frac{1}{\epsilon(x, z)} = \sum_{h=-\infty}^{+\infty} G_h \exp(jh\bar{K} \cdot \bar{r}), \tag{15}$$

where $G_h = \{[(\hat{\epsilon}_0/\epsilon_1)^2 - 1]^{1/2} - (\hat{\epsilon}_0/\epsilon_1)\}^{|h|}/(\hat{\epsilon}_0^2 - \epsilon_1^2)^{1/2}$. Substituting this into the above equation for E_x and using H_y as given by Eq. (12) yields the tangential electric field as

$$E_x = (j/\omega\epsilon_0) \sum_{i=-\infty}^{+\infty} \exp(-j\bar{\sigma}_i \cdot \bar{r}) \times \sum_{h=-\infty}^{+\infty} G_h \left[\frac{dU_{i-h}(z)}{dz} - j(\bar{\sigma}_{i-h} \cdot \hat{z})U_{i-h}(z)\right]. \tag{16}$$

DISCUSSION

The rigorous scalar coupled-wave equations describing diffraction by planar lossy dielectric gratings have been presented for both H-mode and E-mode polarizations as derived from the general vector wave equations. The resulting coupled-wave equations for both cases can be solved in the same manner by using state-variable methods.[2] By using these in combination, any arbitrary input polarization may thus be treated.

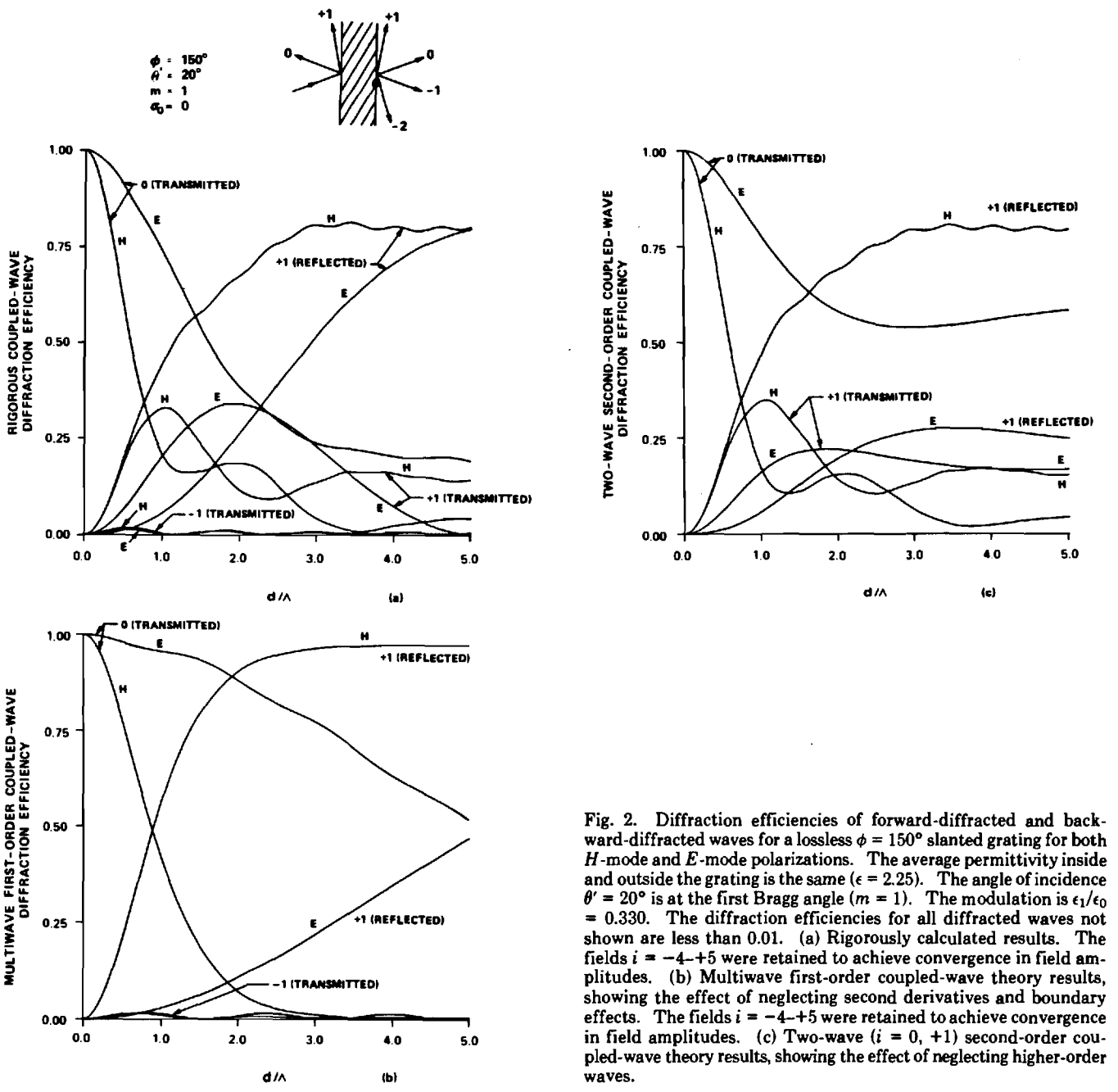


Fig. 2. Diffraction efficiencies of forward-diffracted and backward-diffracted waves for a lossless $\phi = 150°$ slanted grating for both H-mode and E-mode polarizations. The average permittivity inside and outside the grating is the same ($\epsilon = 2.25$). The angle of incidence $\theta' = 20°$ is at the first Bragg angle ($m = 1$). The modulation is $\epsilon_1/\epsilon_0 = 0.330$. The diffraction efficiencies for all diffracted waves not shown are less than 0.01. (a) Rigorously calculated results. The fields $i = -4$–$+5$ were retained to achieve convergence in field amplitudes. (b) Multiwave first-order coupled-wave theory results, showing the effect of neglecting second derivatives and boundary effects. The fields $i = -4$–$+5$ were retained to achieve convergence in field amplitudes. (c) Two-wave ($i = 0, +1$) second-order coupled-wave theory results, showing the effect of neglecting higher-order waves.

For H-mode polarization, the well-known result of direct coupling only between adjacent diffracted orders is obtained. However, for E-mode polarization, it has been shown that direct coupling exists among all diffracted orders. The set of boundary-condition equations for each polarization is a set of linear algebraic equations, and, after the coupled-wave equations are solved, these may then be solved for the phase-matched propagating and evanescent wave amplitudes outside the grating.

Numerous calculations have been performed to obtain the fundamental and higher-order forward and backward-diffracted wave amplitudes for both H-mode-polarization and E-mode-polarization incident waves. Results for an example lossless transmission grating are shown in Fig. 1. The rigorously calculated diffraction efficiencies of several forward-diffracted orders are shown in Fig. 1(a). The power in the E-mode diffracted waves is initially less than that for the H-mode polarization because of the reduced coupling in E mode compared with H mode. However, as the thickness is increased, the E-mode fundamental (+1) diffraction efficiency exceeds the corresponding H-mode diffraction efficiency. Diffraction-efficiency results from multiwave first-order coupled-wave theory[3] are shown for comparison in Fig. 1(b). In this half-space theory, second derivatives of the field amplitudes and boundary effects are neglected. Thus a field that is in fact evanescent (cut off) is still treated as a propagating wave. In Fig. 1, the $i = -1$ field is evanescent. However, this field is predicted by multiwave first-order coupled-wave

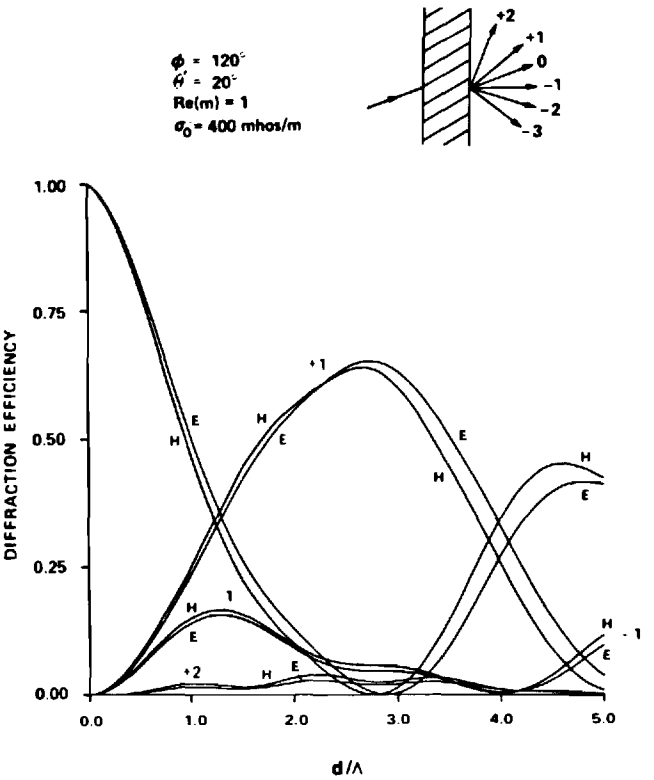


Fig. 3. Rigorously calculated diffraction efficiencies of forward-diffracted waves for a lossy $\phi = 120°$ slanted grating for both H-mode and E-mode polarizations. The average conductivity is $\sigma_0 = 400$ (ohm $- m)^{-1}$. The angle of incidence $\theta' = 20°$ is at the first Bragg angle. The modulation is $\epsilon_1/\epsilon_0 = 0.120$, and the wavelength $\lambda = 514.5$ nm. The average permittivity outside the grating is the same as that inside ($\epsilon = 2.25$).

theory to have a diffraction efficiency as large as 9%, even though this wave is not propagating! Diffraction-efficiency results from two-wave second-order coupled-wave theory[4] are shown for comparison in Fig. 1(c). In this theory, the higher-order waves are neglected, and so only the $i = 0$ and $i = +1$ fields are shown in Fig. 1(c).

Results for an example lossless reflection grating are shown in Fig. 2. The incident wave is at the Bragg angle for $i = +1$ backward-diffracted wave. Rigorously calculated diffraction efficiencies are shown in Fig. 2(a). Diffraction efficiencies from multiwave first-order coupled-wave theory are shown in Fig. 2(b). The poor agreement with rigorously calculated results is apparent and is expected for a reflection grating (see Ref. 1). Diffraction-efficiency results from two-wave second-order coupled-wave theory are presented in Fig. 2(c). The good agreement of the H-mode-polarization results with rigorously calculated results is apparent and is expected for a reflection grating (see Ref. 1). However, for E-mode polarization, the presence of coupling to all higher-order (space-harmonic) fields (not just to adjacent orders, as in the case of H-mode polarization) causes this two-wave theory to give erroneous results for this polarization. The $i = +1$ fundamental backward-diffracted (reflected) wave for E-mode polarization is predicted by this two-wave theory to be much smaller than it actually is. This is a result of artificially restricting the coupling to be between the $i = 0$ and $+1$ space-harmonic fields rather than among all space-harmonic fields.

It may thus be concluded that, to obtain accurate results for E-mode polarization, it is necessary to include higher-order space-harmonic fields regardless of whether the fundamental propagating order is forward or backward diffracted. In addition, if the fundamental propagating order is backward diffracted, the second derivatives and boundary effects need to be included for accurate results for both H-mode and E-mode polarizations.

Rigorously calculated diffraction efficiencies for an example lossy grating are shown in Fig. 3. The nonzero conductivity produces an average absorption that reduces all the diffracted intensities, as would be anticipated.

In separate calculations, it was found that both H-mode and E-mode polarization diffraction efficiencies reduce to the values predicted by Kogelnik's two-wave first-order coupled-wave theory[5] in the limit of sufficiently small modulation.

Coupled-wave analysis is based on the Floquet condition and as such applies to a truly periodic grating (an infinite number of periods). If the grating fringes are exactly parallel to the boundaries ($\phi \equiv 0$), the structure is no longer periodic, and coupled-wave analysis does not apply. In this case, however, a simple rigorous chain-matrix method of analysis may be used.[6]

The generalized rigorous coupled-wave analysis presented here is mathematically exact. There are no theoretical deficiencies or approximations. Any arbitrary level of accuracy is obtainable by increasing the number of orders retained in the analysis. However, convergence is very rapid. In the numerical calculations presented here, the diffracted amplitudes were determined to one part in 10^8. Conservation of power among the beams was accurate to one part in 10^{12}. It should be recognized that this level of accuracy greatly exceeds that usually presented in grating-diffraction calculations.

This research was supported by the National Science Foundation and by the Joint Services Electronics Program.

REFERENCES

1. M. G. Moharam and T. K. Gaylord, "Rigorous coupled-wave analysis of planar-grating diffraction," J. Opt. Soc. Am. 71, 811–818 (1981).
2. C. L. Liu and J. W. S. Liu, *Linear Systems Analysis* (McGraw-Hill, New York, 1975).
3. R. Magnusson and T. K. Gaylord, "Analysis of multiwave diffraction of thick gratings," J. Opt. Soc. Am. 67, 1165–1170 (1977).
4. J. A. Kong, "Second-order coupled-mode equations for spatially periodic media," J. Opt. Soc. Am. 67, 825–829 (1977).
5. H. Kogelnik, "Coupled wave theory for thick hologram gratings," Bell Syst. Tech. J. 48, 2909–2947 (1969).
6. M. G. Moharam and T. K. Gaylord, "Chain-matrix analysis of arbitrary-thickness dielectric reflection gratings," J. Opt. Soc. Am. 72, 187–190 (1982).

Appl. Phys. B 28, 1–14 (1982)

Applied Physics B Photophysics and Laser Chemistry

Planar Dielectric Grating Diffraction Theories

T. K. Gaylord and M. G. Moharam

School of Electrical Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA

Received 23 December 1981/Accepted 26 February 1982

Abstract. Various planar dielectric grating diffraction theories are reviewed for the case of a general sinusoidal permittivity planar grating with slanted fringes and plane wave incidence at an arbitrary angle. Exact formulations without approximations (rigorous coupled-wave analysis and rigorous modal analysis) are developed first. Then, using a series of fundamental assumptions, rigorous theory is shown to reduce to the various approximate theories in the appropriate limits. The implications of these fundamental assumptions are discussed.

PACS: 42.10, 42.20, 42.30

Since 1930 there have been over 400 scientific papers on the subject of grating diffraction. Many of these papers have been applicable to planar dielectric gratings. These periodic structures have been applied in numerous areas such as acousto-optics, holography, integrated optics, and spectral analysis. The diffraction of electromagnetic waves by spatially periodic media may be analyzed by numerous methods and with a wide variety of possible assumptions. The purpose of this paper is to review both rigorous and approximate planar grating diffraction theories and to show explicitly the relationships between the various theories.

The most common methods of analyzing planar dielectric grating diffraction are the coupled-wave approach [1–8] and the modal approach [9–18]. These theories have recently been treated in two extensive reviews [19, 20]. Both coupled-wave and modal approaches can produce exact formulations without approximations. In their full rigorous forms these formulations are completely equivalent [21]. They represent merely alternative methods of representing the electromagnetic fields inside the grating (Sect. 2).

Starting with the rigorous theories and using a series of fundamental assumptions, these general theories are shown to reduce to the various approximate theories [two-wave modal theory, two-wave second-order coupled-wave theory, multiwave coupled-wave theory, two-wave first-order coupled-wave theory (Kogelnik theory), Raman-Nath theory, and amplitude transmittance theory] in the appropriate limits. This is shown in Sect. 8.

1. Planar Dielectric Grating Diffraction

The general planar grating diffraction problem is depicted in Fig. 1. An electromagnetic wave is obliquely incident upon a slanted-fringe planar grating bounded by two different homogeneous media. In general, there will be simultaneously both forward-diffracted and backward-diffracted waves as shown in the figure. This geometry is applicable 1) to holographic gratings in air or other media ($\varepsilon_1 = \varepsilon_{III} \neq \varepsilon_0$), 2) to acousto-optic gratings within a medium ($\varepsilon_1 = \varepsilon_0 = \varepsilon_{III}$), and 3) to grating couplers such as used in integrated optics ($\varepsilon_1 \neq \varepsilon_0 \neq \varepsilon_{III} \neq \varepsilon_I$). The quantities ε_I, ε_0, and ε_{III} are the average relative permittivities (dielectric constants) in regions 1, 2, and 3, respectively.

In this paper, for simplicity, the case of a lossless dielectric grating with sinusoidal permittivity is treated. The incident plane wave polarization is perpendicular to the plane of incidence (H mode). This is probably the case of widest general interest. However, these assumptions are not essential to the theories described. The relative permittivity (dielectric constant) in the

0721-7269/82/0028/0001/$02.80

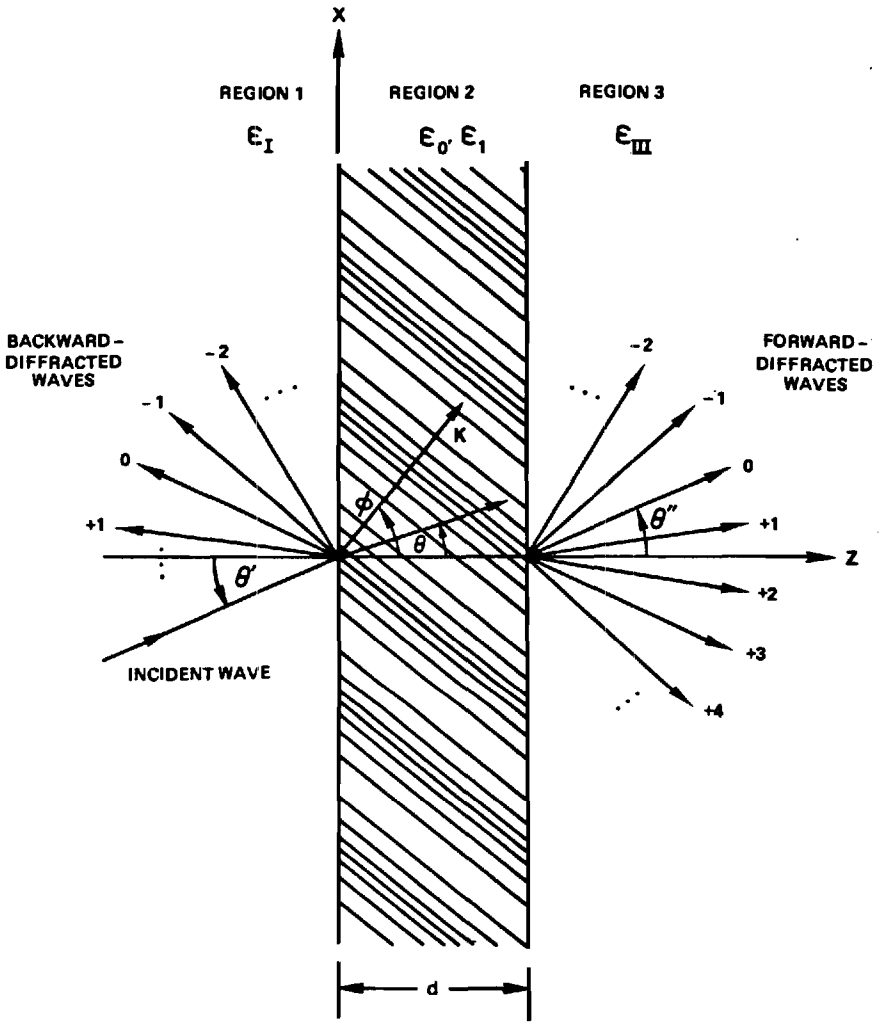


Fig. 1. Geometrical configuration of planar grating diffraction

grating region is given by

$$\varepsilon(x,z)=\varepsilon_0+\varepsilon_1\cos(\bar{K}\cdot\bar{r})$$
$$=\varepsilon_0+\varepsilon_1\cos[K(x\sin\phi+z\cos\phi)], \quad (1)$$

where ε_1 is the amplitude of the sinusoidal relative permittivity, ϕ is the grating slant angle, and $K=2\pi/\Lambda$, where Λ is the grating period. The cosinusoidal form used in (1) is common in the volume holographic grating literature. In the acousto-optics literature, a sinusoidal form for (1) is more common. Using the sinusoidal form would alter the resulting equations in the following sections as well as their amplitude solutions. However, the diffracted intensities are identical in either case.

The general approach to the planar-grating diffraction problem involves finding a solution of the wave equation in each of the three regions and then matching the tangential electric and magnetic fields at the two interfaces ($z=0$ and $z=d$). In region 1, the normalized amplitude of the incident plane wave is

$$E_{inc}=\exp(-j\bar{k}_1\cdot\bar{r})$$
$$=\exp[-jk_1(x\sin\theta'+z\cos\theta')], \quad (2)$$

where θ' is the angle of incidence on region 1, $k_1=2\pi(\varepsilon_1)^{1/2}/\lambda$, and λ is the free space wavelength. The wave equation for the H mode polarization is the scalar wave equation (Helmholtz equation)

$$\nabla^2E+k^2\varepsilon(x,z)E=0, \quad (3)$$

where $k=2\pi/\lambda$. For H mode polarization, the electric field only has a component in the y direction. The fields and the grating are unchanging in the y direction. For any arbitrary direction the grating as bounded by regions 1 and 3 is periodic only in the x component of the direction. If region 2 was infinite in all directions (not bounded) the resulting grating would be periodic in any direction that was not perpendicular to the grating vector $\bar{K}$. The Floquet theorem [22, 23] restricts the possible fields that can exist in a periodic structure at steady state. As a result of the Floquet theorem the diffracted wavevectors inside the grating may be represented for the infinite periodic medium case by

$$\bar{\sigma}_i=\bar{k}_2-i\bar{K}, \quad (4)$$

where $\bar{\sigma}_i$ is the wavevector of the i-th space harmonic in the grating, i is any integer, and $\bar{k}_2$ is the wavevector of the zero-order ($i=0$) space harmonic having a magnitude of $k_2=2\pi(\varepsilon_0)^{1/2}/\lambda$. For the bounded grating that is periodic only in the x-component of direction, the Floquet theorem only requires

$$\bar{\sigma}_i\cdot\hat{x}=(\bar{k}_2-i\bar{K})\cdot\hat{x}, \tag{5}$$

where $\hat{}$ denotes a unit vector. This expression is just the x-component of (4). Only the form as represented by (5) is necessary for the present problem. However, for subtle reasons that will become clear in the next section, the Floquet requirement as expressed by (4) will be used. This is certainly acceptable since it contains the necessary (5) within it.

2. Rigorous Coupled-Wave and Modal Theories

It is possible to formulate the planar grating diffraction problem depicted in Fig. 1 in an exact manner. This may be done with the coupled-wave approach or the modal approach. The modal approach is sometimes referred to as the Floquet, Floquet-Bloch, eigenmode, characteristic-mode, or coupled-mode approach. The coupled-wave approach is confusingly also sometimes called coupled-mode approach. Both the coupled-wave and modal approaches are altenrative methods of representing the fields inside the grating medium.

In the coupled-wave representation, the fields inside the grating are expanded in terms of the space harmonics of the fields in the periodic structure. These space harmonics inside the grating correspond to diffracted orders outside of the grating. Thus, the partial fields inside the modulated medium are visualized as diffracted waves that progress through the planar slab and couple energy back and forth between each other as they progress. This picture agrees rather well with simple physical intuition about the process of diffraction by a volume grating. In the coupled-wave approach the total field is thus expressed as

$$E(x,z)=\sum_{i=-\infty}^{+\infty} S_i(z)\exp(-j\bar{\sigma}_i\cdot\bar{r}), \tag{6}$$

where i is the space harmonic index. Equation (6) has the general appearance of a plane wave expansion of diffracted waves with amplitudes S_i. This would be true if the S_i's were constants. However, since the S_i's are not constants but are functions of z, each i does not correspond to a single plane wave. In general there are an infinite number of plane waves associated with each i. S_i varies only in the direction perpendicular to the boundary. The sum of all of the i-th partial fields as represented by $E(x,z)$ in (6) satisfies the wave equation. However, individually the partial fields do not satisfy the wave equation.

In the modal representation, the fields inside the grating are expanded in terms of the allowable modes of the periodic medium. The fields are visualized as waveguide modes in the grating region. In the modal approach, the total electric field is expressed as a weighted summation over all possible modes,

$$E(x,z)=\sum_{m=-\infty}^{+\infty} C_m\Phi_m(\bar{r})\exp(-j\bar{k}_{2m}\cdot\bar{r}), \tag{7}$$

where m is the mode index. The function $\Phi_m(\bar{r})$ is periodic with a period equal to the grating period. That is $\Phi_m(\bar{r})=\Phi_m(\bar{r}+\bar{\Lambda})$. The summation includes both forward and backward propagating modes. The backward propagating modes are due to diffraction in the grating volume (when the grating fringes are slanted) and due to reflections at the $z=d$ boundary. Each individual m-th mode satisfies the wave equation and may be either evanescent or propagating. These modes in the grating are precisely analogous to modes in a waveguide. Each waveguide mode satisfies the wave equation by itself and it may be either cutoff or propagating. Each mode (m) consists of an infinite number of space harmonics (i) and each mode propagates through the medium without change. The space harmonics may be viewed as arising from the Fourier expansion of the periodic function $\Phi_m(\bar{r})$.

The coupled-wave representation or expansion in terms of space harmonics, (6), and the modal representation or expansion in terms of modes, (7), are merely alternative representations of the same physical problem. Both approaches are complete and both are rigorous formulations (without approximations). These two approaches are thus completely equivalent and this will be demonstrated in Sect. 5.

3. Rigorous Coupled-Wave Equations

The rigorous coupled-wave equations are developed by starting with the field expansion in terms of space harmonics, (6). There are mutliple versions of the rigorous coupled-wave equations for the physical situation depicted in Fig. 1 depending on the form of the modulation chosen (sinusoidal or cosinusoidal permittivity), the form of the Floquet condition chosen, (4) or (5), and the polarization chosen (H mode or E mode). The basic case treated in this paper is cosinusoidal permittivity and H mode polarization. Both forms of the Floquet condition are analyzed in this section.

The Floquet condition for an infinite periodic medium, (4), will be treated first since this gives rise to the very useful, common form of the rigorous coupled-wave equations. In fact, only the component of the wavevectors along the boundary in the direction of periodicity

(x) needs to satisfy the Floquet condition. This is given by (5) and will be treated below. The Floquet condition for an infinite grating, (4), contains the required x dependence. The z dependence in (4) will be shown to be very useful in obtaining a directly solvable form of the coupled-wave equations. Substituting (4) into (6) yields

$$E(x,z)=\sum_{i=-\infty}^{+\infty} S_i(z)\exp[-j(\bar{k}_2-i\bar{K})\cdot\bar{r}]$$
$$=\sum_{i=-\infty}^{+\infty} S_i(z)\exp\{-j[(k_2\sin\theta-iK\sin\phi)x+(k_2\cos\theta-iK\cos\phi)z]\}, \quad (8)$$

where θ is the angle of refraction of the incident beam from region 1. Thus θ is related to θ' through

$$k_1\sin\theta'=k_2\sin\theta. \quad (9)$$

Substituting (1) and (8) into (3) and performing the indicated differentiations gives

$$\sum_{i=-\infty}^{+\infty}\left\{\frac{\partial^2 S_i(z)}{\partial z^2}-j2(k_2\cos\theta-iK\cos\phi)\frac{\partial S_i(z)}{\partial z}-[(k_2\sin\theta-iK\sin\phi)^2+(k_2\cos\theta-iK\cos\phi)^2]S_i(z)+k^2\varepsilon_0 S_i(z)+\frac{k^2\varepsilon_1}{2}S_{i-1}(z)+\frac{k^2\varepsilon_1}{2}S_{i+1}(z)\right\}\cdot\exp\{-j[(k_2\sin\theta-iK\sin\phi)x+(k_2\cos\theta-iK\cos\phi)z]\}=0. \quad (10)$$

This equation must be satisfied for all values of the variables. Thus the coefficient of each exponential must individually be zero for nontrivial solutions. Using this and the definitions of k, k_2, and K, (10) reduces to the rigorous coupled-wave equations:

$$\frac{1}{2\pi^2}\frac{d^2S_i(z)}{dz^2}-j\frac{2}{\pi}\left[\frac{(\varepsilon_0)^{1/2}\cos\theta}{\lambda}-\frac{i\cos\phi}{\Lambda}\right]\frac{dS_i(z)}{dz}+\frac{2i(m-i)}{\Lambda^2}S_i(z)+\frac{\varepsilon_1}{\lambda^2}[S_{i+1}(z)+S_{i-1}(z)]=0. \quad (11)$$

This is an infinite set of second-order coupled difference-differential equations. By inspection, it is seen that the wave corresponding to each value of i (space harmonic inside the grating or diffracted order outside of the grating) is coupled to its adjacent ($i+1$ and $i-1$) space harmonics. There is no direct coupling between nonadjacent orders.

In the rigorous coupled-wave equations the quantity m has been defined as

$$m=\frac{2\Lambda(\varepsilon_0)^{1/2}}{\lambda}\cos(\theta-\phi). \quad (12)$$

The quantity m may have any value in general. For the case when m is an integer, (12) becomes the Bragg condition. However, it is important to realize that the Bragg condition is not specifically an input into this theory. The approach applies to an arbitrary angle of incidence and wavelength. Only if the angle of incidence and wavelength are such that m is an integer does Bragg incidence occur. The rigorous coupled-wave equations given by (11) are in a from that is directly solvable using a state variables approach from linear systems theory. This method of solution will be used in the next section.

The Floquet condition as given by (5) may also be used in space harmonic field expansion, (6), to obtain an alternative set of rigorous coupled-wave equations. Substituting (5) into (6) yields

$$E(x,z)=\sum_{i=-\infty}^{+\infty} S_i(z)\exp[-j(k_{2x}-iK_x)x]$$
$$=\sum_{i=-\infty}^{+\infty} S_i(z)\exp[-j(k_2\sin\theta-iK\sin\phi)x]. \quad (13)$$

The field expansion given by (13) is the same as that in (8) except that the z dependent part of the exponential has been included in the $S_i(z)$ functions in (13). Substituting (1) and (13) into (3), performing the indicated differentiations, and setting the coefficients of each exponential equal to zero as before, produces

$$\frac{1}{2\pi^2}\frac{d^2S_i(z)}{dz^2}-2\left\{\left[\frac{(\varepsilon_0)^{1/2}\sin\theta}{\lambda}-\frac{i\sin\phi}{\Lambda}\right]^2-\frac{\varepsilon_0}{\lambda^2}\right\}S_i(z)+\frac{\varepsilon_1}{\lambda^2}\exp\left(+j\frac{2\pi z\cos\phi}{\Lambda}\right)S_{i-1}(z)+\frac{\varepsilon_1}{\lambda^2}\exp\left(-j\frac{2\pi z\cos\phi}{\Lambda}\right)S_{i+1}(z)=0. \quad (14)$$

This set of coupled-wave equations contains no first derivative terms in contrast to the coupled-wave equations given in (11). In addition, (14) is a nonconstant coefficient differential equation due to the presence of z in the coefficients of the $S_{i-1}(z)$ and $S_{i+1}(z)$ terms. The equations in the form of (14) represent a linear shift-variant system and direct solution would be difficult. For the case of an unslanted grating ($\phi=\pi/2$, fringes perpendicular to the surface), the equations become constant coefficient differential equations. For this limiting case, the equations become identical to the coupled-wave equations of Kong [Ref. 6, Eqs. (6a) and (6b)] if only two waves are retained ($i=0, 1$).

4. Solution of the Rigorous Coupled-Wave Equations

The rigorous coupled-wave equations as given by (11) represent a set of second-order linear differential equations with constant coefficients. Using the methods of linear systems analysis [24] this differential equation description of this continuous system may be trans-

formed into a state space description and a solution obtained directly. By defining the state variables as

$$S_{1,i}(z)=S_i(z), \tag{15}$$

$$S_{2,i}(z)=\frac{dS_i(z)}{dz} \tag{16}$$

the infinite set of second-order differential equations (11) are transformed into two infinite sets of first-order differential equations:

$$\frac{dS_{1,i}(z)}{dz}=S_{2,i}(z), \tag{17}$$

$$\begin{aligned}\frac{dS_{2,i}(z)}{dz}=&-\frac{2\pi^2\varepsilon_1}{\lambda^2}S_{1,i-1}(z)+\frac{4\pi^2 i(i-m)}{\Lambda^2}S_{1,i}(z)\\&-\frac{2\pi^2\varepsilon_1}{\lambda^2}S_{1,i+1}(z)\\&+j4\pi\left(\frac{(\varepsilon_0)^{1/2}\cos\theta}{\lambda}-\frac{i\cos\phi}{\Lambda}\right)S_{2,i}(z).\end{aligned} \tag{18}$$

Equations (17) and (18) are the state equations corresponding to the rigorous coupled-wave equations (11). Since these are homogeneous equations, they correspond to unforced state equations. State equations that are linear differential equations with constant coefficients such as these, may be solved for closed-form expressions for the state variables. In this case, only the homogeneous solution is necessary as there are no driving terms in these equations. The homogeneous solutions are

$$S_{l,i}(z)=\sum_{m=-\infty}^{+\infty}C_m w_{l,im}\exp(\lambda_m z) \tag{19}$$

for $l=1,2$. The coefficients C_m are unknown constants to be determined from the boundary conditions. The quantity $w_{l,im}$ is an element of an eigenvector and λ_m is an eigenvalue. These quantities are determined as described below. The solution for the wave amplitudes (the "output equation" in linear systems terminology) is $S_i(z)=S_{1,i}(z)$.

The constituent state equations (17) and (18) may be written in matrix form as

$$\begin{bmatrix}\vdots\\ \dot S_{1,-2}\\ \dot S_{1,-1}\\ \dot S_{1,0}\\ \dot S_{1,1}\\ \dot S_{1,2}\\ \vdots\\ \dot S_{2,-2}\\ \dot S_{2,-1}\\ \dot S_{2,0}\\ \dot S_{2,1}\\ \dot S_{2,2}\\ \vdots\end{bmatrix}=\begin{bmatrix}&&&\vdots&&&&&\vdots&&&\\ 0&0&0&0&0&1&0&0&0&0&\\ 0&0&0&0&0&0&1&0&0&0&\\ \cdots&0&0&0&0&0&\cdots&0&0&1&0&0&\cdots\\ 0&0&0&0&0&0&0&0&1&0&\\ 0&0&0&0&0&0&0&0&0&1&\\ &&&\vdots&&&&&\vdots&&&\\ b_{-2}&a&0&0&0&c_{-2}&0&0&0&0&\\ a&b_{-1}&a&0&0&0&c_{-1}&0&0&0&\\ \cdots&0&a&b_0&a&0&\cdots&0&0&c_0&0&0&\cdots\\ 0&0&a&b_1&a&0&0&0&c_1&0&\\ 0&0&0&a&b_2&0&0&0&0&c_2&\\ &&&\vdots&&&&&\vdots&&&\end{bmatrix}\begin{bmatrix}\vdots\\ S_{1,-2}\\ S_{1,-1}\\ S_{1,0}\\ S_{1,1}\\ S_{1,2}\\ \vdots\\ S_{2,-2}\\ S_{2,-1}\\ S_{2,0}\\ S_{2,1}\\ S_{2,2}\\ \vdots\end{bmatrix}, \tag{20}$$

where $a=-2\pi^2\varepsilon_1/\lambda^2$, $b_i=4\pi^2 i(i-m)/\Lambda^2$, and $c_i=j4\pi[(\varepsilon_0)^{1/2}\cos\theta/\lambda-i\cos\phi/\Lambda]$. This equation may be represented concisely as $\dot{\mathbf{S}}=\mathbf{AS}$ where $\dot{\mathbf{S}}$ and $\mathbf{S}$ are the column vectors in (20) and $\mathbf{A}$ is the coefficient matrix. The needed eigenvalues and eigenvectors are determined from this coefficient matrix. Although $\mathbf{A}$ is an infinite matrix, results may be obtained in practice to an arbitrary level of accuracy with a truncated matrix. Each of the four submatrices is truncated to $n\times n$. As the integer n increases, the calculated results rapidly converge to the exact results. The quantity n corresponds to the total number of space harmonics retained in the analysis. This in turn means that the analysis includes n diffracted waves in region 1 and n diffracted waves in region 3. To put the four submatrices into standard form, the integers i and m are replaced with the new integers p and q that run from 1 to n. For example, if an odd number of waves are retained symmetrically about $i=0$ (the undiffracted transmitted wave) in the analysis, then $p=i+(n+1)/2$ and $q=m+(n+1)/2$. The $2n$ solutions may then be expressed

$$S_{l,p}(z)=\sum_{r=1}^{2}\sum_{q=1}^{n}C_{r,q}w_{l,p;r,q}\exp(\lambda_{r,q}z) \tag{21}$$

for $l=1,2$ and $p=1$ to n. The eigenvalues $\lambda_{r,q}$ are determined by solving the determinantal equation

$$|\mathbf{A}-\lambda_{r,q}\mathbf{I}|=0, \tag{22}$$

where $\mathbf{I}$ is the unit matrix. The eigenvector corresponding to a particular eigenvalue $\lambda_{r,q}$ is determined by substituting $2n$ expressions ($l=1,2$ and $p=1$ to n) for $S_{l,p}$ of the form $S_{l,p}=B_{l,p;r,q}\exp(\lambda_{r,q})$ into the state equation (20), performing the indicated differentiations, and then solving for each element of the eigenvector as $w_{l,p;r,q}=B_{l,p;r,q}/B_{1,1\pi;r,q}$ using Cramer's rule and thus expressing each element as a ratio of determinants. The eigenvalues and eigenvectors for a

matrix are typically calculated numerically using a computer library program [25].

5. Equivalence of Coupled-Wave and Modal Representations

The total field inside the grating may be expressed in coupled-wave form, (6), or in modal form, (7), depending whether the field is expanded in terms of space harmonics or in terms of modes, respectively. These two forms are alternative representations that are completely equivalent. This equivalence has been discussed previously [14, 19, 21]. It can be shown mathematically in a simple manner as follows.

Substituting $S_i(z) = \sum_{m=-\infty}^{+\infty} C_m w_{1,im} \exp(\lambda_m z)$ into the coupled-wave expansion (8) gives

$$E(x,z) = \sum_{i=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} C_m w_{1,im} \cdot \exp\{-j[(k^2 \sin\theta - iK\sin\phi)x + (k_2\cos\theta - iK\cos\phi + j\lambda_m)z]\}. \quad (23)$$

Changing the order of the summation this may be rewritten as

$$E(x,z) = \sum_{m=-\infty}^{+\infty} C_m \sum_{i=-\infty}^{+\infty} w_{1,im} \exp[-j(\bar{k}_{2m} - i\bar{K})\cdot\bar{r}], \quad (24)$$

where $\bar{k}_{2m} = k_2 \sin\theta\,\hat{x} + (k_2\cos\theta + j\lambda_m)\hat{z}$. Identifying the complex Fourier series and its representation of a periodic function

$$\sum_{i=-\infty}^{+\infty} w_{1,im} \exp(ji\bar{K}\cdot\bar{r}) = \Phi_m(\bar{r}) = \Phi_m(\bar{r} + \bar{\Lambda}) \quad (25)$$

gives the modal expansion

$$E(x,z) = \sum_{m=-\infty}^{+\infty} C_m \Phi_m(\bar{r}) \exp(-j\bar{k}_{2m}\cdot\bar{r}) \quad (26)$$

which is identical to (7). Thus the coupled-wave and modal representations are seen to be equivalent.

6. Phase Matching and Boundary Conditions

Each i-th field in region 1 and 3 must be phase matched to the i-th space harmonic field inside the grating. In addition, the magnitude of the fields in regions 1 and 3 must be such that the electromagnetic boundary conditions are satisfied at the two grating boundaries ($z=0$ and $z=d$).

The total electric field in region 1 is the sum of the incident and the backward-traveling waves. The normalized total electric field in region 1 may be expressed as

$$E_1 = \exp(-j\bar{k}_1\cdot\bar{r}) + \sum_{i=-\infty}^{\infty} R_i \exp(-j\bar{k}_{1i}\cdot\bar{r}), \quad (27)$$

where R_i is the normalized amplitude of the i-th reflected wave in region 1 with wavevector $\bar{k}_{1i}$. Likewise the normalized total electric field in region 3 is

$$E_3 = \sum_{i=-\infty}^{\infty} T_i \exp[-j\bar{k}_{3i}\cdot(\bar{r} - d\hat{z})], \quad (28)$$

where T_i is the normalized amplitude of the i-th transmitted wave in region 3 with wavevector $\bar{k}_{3i}$. These fields in regions 1 and 3 are phased matched to the field in the grating, (8). Thus the x components of the wavevectors of the i-th wave (regions 1 and 3) and the x component of the wavevector of the i-th space harmonic field (region 2) must be the same. That is

$$\bar{k}_{1i}\cdot\hat{x} = (\bar{k}_2 - i\bar{K})\cdot\hat{x} = \bar{k}_{3i}\cdot\hat{x}$$
$$k_1\sin\theta_i' = k_2\sin\theta - iK\sin\phi = k_3\sin\theta_i''. \quad (29)$$

In the homogeneous regions (1 and 3) the backward- and forward-diffracted waves have wavevectors with magnitudes

$$|\bar{k}_{1i}| = |\bar{k}_1| \quad \text{and} \quad |\bar{k}_{3i}| = |\bar{k}_3|, \quad (30)$$

where $k_3 = 2\pi(\varepsilon_{III})^{1/2}/\lambda$. Knowing the total amplitudes and the x components of the diffracted wavevectors, the z components are then determined to be

$$\bar{k}_{1i}\cdot\hat{z} = [|\bar{k}_1|^2 - (\bar{k}_{1i}\cdot\hat{x})^2]^{1/2} = [k_1^2 - (k_2\sin\theta - iK\sin\phi)^2]^{1/2} \quad (31)$$

and

$$\bar{k}_{3i}\cdot\hat{z} = [|\bar{k}_3|^2 - (\bar{k}_{3i}\cdot\hat{x})^2]^{1/2} = [k_3^2 - (k_2\sin\theta - iK\sin\phi)^2]^{1/2}. \quad (32)$$

These quantities are either positive real (propagating wave) or negative imaginary (evanescent wave).

Using the phase matching conditions (29), (31), and (32), the total fields in regions 1 and 3, (27) and (28), may be rewritten

$$E_1 = \exp\{-j[k_1(\sin\theta' x + \cos\theta' z)]\} + \sum_{i=-\infty}^{\infty} R_i \exp[-j\{(k_2\sin\theta - iK\sin\phi)x - [k_1^2 - (k_2\sin\theta - iK\sin\phi)^2]^{1/2} z\}] \quad (33)$$

and

$$E_3 = \sum_{i=-\infty}^{\infty} T_i \exp\{-j\{(k_2\sin\theta - iK\sin\phi)x + [k_3^2 - (k_2\sin\theta - iK\sin\phi)^2]^{1/2}(z-d)\}\}. \quad (34)$$

Electromagnetic boundary conditions require that the tangential electric and tangential magnetic fields be continuous across the two boundaries ($z=0$ and $z=d$). For the H mode polarization described in this paper, the electric field only has a component in the y

direction and so it is the tangential electric field directly. The magnetic field intensity, however, must be obtained through the Maxwell equation $\nabla x\bar{E} = -\partial\bar{B}/\partial t$. The tangential component of H is in the x direction and is thus given by $H_x = (-j/\omega\mu)\partial E_y/\partial z$. For each value of i, the four quantities to be matched and the resulting boundary condition are:

1) tangential E at $z=0$:

$$\delta_{i0} + R_i = S_i(0), \tag{35}$$

2) tangential H at $z=0$:

$$j[k_1^2 - (k_2 \sin\theta - iK \sin\phi)^2]^{1/2}(R_i - \delta_{i0}) = \frac{dS_i(0)}{dz}, \tag{36}$$

3) tangential E at $z=d$:

$$T_i = S_i(d) \exp[-j(k_2 \cos\theta - iK \cos\phi)d], \tag{37}$$

4) tangential H at $z=d$:

$$-j[k_3^2 - (k_2 \sin\theta - iK \sin\phi)^2]^{1/2} T_i = \left[\frac{dS_i(d)}{dz} - j(k_2 \cos\theta - iK \cos\phi) S_i(d)\right] \cdot \exp[-j(k_2 \cos\theta - iK \cos\phi)d], \tag{38}$$

where δ_{i0} is the Kronecker delta function.

7. Diffraction Efficiency

The quantity commonly measured in grating diffraction is the diffraction efficiency. It is defined as the diffracted intensity of a particular order divided by the input intensity. In the above formulation, the incident plane wave amplitude was normalized to unity. Thus the diffraction efficiencies in regions 1 and 3 are

$$\begin{aligned} \mathrm{DE}_{1i} &= \mathrm{Re}\{(\bar{k}_{1i} \cdot \hat{z})/(\bar{k}_{10} \cdot \hat{z})\} R_i R_i^* \\ &= \mathrm{Re}\{\{1 - [\sin\theta' - i\lambda \sin\phi/(\varepsilon_1)^{1/2} \Lambda]^2\}^{1/2}/ \\ &\quad \cos\theta'\} R_i R_i^* \end{aligned} \tag{39}$$

and

$$\begin{aligned} \mathrm{DE}_{3i} &= \mathrm{Re}\{(\bar{k}_{3i} \cdot \hat{z})/(\bar{k}_{10} \cdot \hat{z})\} T_i T_i^* \\ &= \mathrm{Re}\{\{(\varepsilon_{III}/\varepsilon_1) - [\sin\theta' - i\lambda \sin\phi/(\varepsilon_1)^{1/2} \Lambda]^2\}^{1/2}/ \\ &\quad \cos\theta'\} T_i T_i^* . \end{aligned} \tag{40}$$

The real part of the ratio of the propagation constants occurs when the time-average power-flow density is obtained by taking the real part of the complex Poynting vector. For an unslanted grating ($\phi = \pi/2$) with the same medium on both sides ($\varepsilon_1 = \varepsilon_{III}$), the real part of the ratio of the propagation constants is just the usual ratio of the cosine of the diffraction angle for the i-th diffracted wave to the cosine of the incidence angle.

If n values of i are retained in the analysis, then there will be n forward-diffracted waves (n values of T_i) and n backward-diffracted waves (n values of R_i). Correspondingly, there will be $2n$ unknown values of C_m. This is because the coefficient matrix in (20) is a $2n \times 2n$ matrix and therefore has $2n$ eigenvalues and thus there are $2n$ unknown values of C_m. Also this may be viewed as being due to the n coupled-wave equations, each being a second-order differential equation, and thus there are $2n$ roots or eigenvalues and $2n$ unknown constants C_m to be determined from the boundary conditions. Therefore, the total number of unknowns is $4n$. Substituting $S_i(z)$, as given by (15) and (19), into the equations for the boundary conditions (35)–(38) produces n linear equations containing the $4n$ unknowns. An efficient procedure to solve these equations is to eliminate R_i and T_i from these equations and to solve the resulting $2n$ equations for the $2n$ values of C_m using a technique such as Gauss elimination. Then the n values of R_i and n values of T_i may be determined from (35) and (37) respectively. Finally, the diffraction efficiencies DE_{1i} and DE_{3i} are calculated using (39) and (40). For phase gratings the input power is conserved and thus the sum of all of the efficiencies for the propagating waves is unity. That is,

$$\sum_i (\mathrm{DE}_{1i} + \mathrm{DE}_{3i}) = 1. \tag{41}$$

Equation (41) may be used to verify the convergence of the numerical calculations.

8. Approximate Theories

The vast majority of the papers on grating diffraction theory have dealt with approximate theories. There are a large number of possible approximations and assumptions that can be made. These generally lead to enormous simplifications in the analyses. In some cases, these simplifications allow analytic solutions to be obtained. A number of famous analytic expressions occur for special limiting cases.

In this section, a large number of planar grating diffraction theories are classified in terms of the fundamental assumptions: 1) neglect of higher-order waves, 2) neglect of second derivatives of the field amplitudes, 3) neglect of boundary effects, 4) neglect of dephasing from the Bragg condition, and 5) the small grating modulation approximation. In addition to these assumptions, a number of other approximations such as normal incidence, unslanted gratings, and large grating period compared to a wavelength, may also be made. However, in this section, only the fundamental assumptions enumerated above are treated. Thus all of the approximate theories are presented in their general form allowing for arbitrary angle of incidence (θ'),

arbitrary grating slant angle (ϕ), and arbitrary grating period (Λ). The various further reductions can then be easily formulated, if desired, from these general forms of the approximate theories.

In region 1 of Fig. 1, backward-traveling waves exist. In general, these waves are produced both by diffraction from within the grating volume and by boundary effects (diffraction and reflection from the periodic boundaries at $z=0$ and $z=d$). These physical processes produce a spectrum of plane waves traveling back into region 1 ($z<0$). For the general planar grating of Fig. 1, neglecting the second derivatives of the field amplitudes in the wave equation reduces the number of waves in the analysis from $2n$ to n. The bulk diffracted orders are retained and the boundary-produced waves are eliminated. Thus for a planar grating, the neglect of second derivatives and the neglect of boundary effects are absolutely linked together. When these assumptions are made, the resulting first-order coupled-wave analyses have the amplitudes of the diffracted waves calculated *inside* the modulated region. Then the amplitudes T_i of the forward-diffracted output waves are obtained (approximately) by arguing that they are equal to $S_i(d)$, the space harmonic field amplitude at a distance d from the input surface $z=0$. Likewise for those values of i that represent backward-diffracted waves, the amplitudes R_i are estimated to be $S_i(0)$. However, in the physical problem being analyzed, there are no boundaries at $z=0$ and $z=d$. These planes just represent reference locations. There are no reflected or diffracted waves resulting from these planes and thus there are no physical boundaries at these locations! Thus, the assumptions of neglecting the second derivatives of field amplitudes and neglecting boundary effects have transformed the problem into a filled-space problem (a grating filling all space) with imaginary boundaries at $z=0$ and $z=d$ that are used only to obtain an approximate mathematical formulation of the problem. The first-order theory approaches are not capable of solving the problem of the general planar slab grating bounded by two media different from the grating medium. These two linked assumptions therefore, unmistakably imply the filled-space problem. *After* the filled-space problem is solved, *then* it is assumed that the grating terminates at $z=0$ and $z=d$ and, as a result, that $T_i \cong S_i(d)$ for the forward-diffracted waves and $R_i \cong S_i(0)$ for the backward-diffracted waves. This is obviously only an approximation to the actual situation.

Another consequence of neglecting second derivatives is the exclusion of some propagating waves. In first-order theory, only half of the waves can be retained in the analysis. That is, only one set of i values (as opposed to two sets) is included. For a genral slanted grating, some of these waves may be forward-diffracted and some of them may be backward-diffracted. From (8), if $k_z \cos\theta - iK\cos\phi$ is positive, the wave is forward-diffracted and if negative, it is backward-diffracted. For forward-diffracted waves, the boundary condition used must be $S_i(0)=0$. For backward-diffracted waves, the appropriate boundary condition is $S_i(d)=0$. The second set of waves (set of i values) are phase matched to these waves. This second set of waves is, of course, neglected in any first order analysis. For the example depicted in Fig. 2, the backward-diffracted waves for $-1 \leqq i \leqq +4$ would all be neglected in first-order theory. The diffraction efficiencies of these backward-diffracted waves are arbitrarily set equal to zero. For the case of a slanted-fringe grating, the power in the neglected phase-matched waves has been shown to be very significant in some cases [8]. Thus the errors introduced by using first order theory can be particularly significant for slant angles away from $\phi=0$ and $\phi=\pi/2$.

Still another consequence of neglecting second derivatives is the exclusion of evanescent waves from the analysis. In first order theory, the filled-space nature of the grating being analyzed, causes one complete set of diffracted orders (i) to exist inside the grating, since all of the $S_i(z)$'s exist there. These calculated values of $S_i(z)$ may have wavevectors with components either in the $+$ or $-$ z directions. However, many of the wavevectors of the $S_i(z)$'s cannot be phased matched to plane waves outside of the grating (regions 1 and 3). This may be seen from Fig. 2. For this example, the values $-1 \leqq i \leqq +4$ correspond to propagating plane waves in regions 1 and 3. The values $i \leqq -2$ and $i \geqq +5$ correspond to evanescent waves in region 1 and 3. However, in first order analysis (without second derivatives) all values of i are treated as representing propagating waves. This is obviously not true. Nevertheless, diffraction efficiencies can be calculated for these evanescent waves as though they were propagating. These predicted efficiencies are clearly incorrect since they should be zero. If the grating period is much larger than a wavelength ($\Lambda \gg \lambda$), then there will be a large number of propagating waves and the effect of excluding evanescent waves would be reduced.

Therefore, it is concluded that all first-order theories inherently contain: 1) the approximate method for calculating diffracted amplitudes described above, 2) neglect of phase-matched waves, and 3) neglect of evanescent waves.

A depiction of various planar grating diffraction theories and their interrelationships in terms of fundamental assumptions is shown in Fig. 3. Most of the literature on planar grating diffraction theory can be connected with a particular block in this diagram. The importance of the various assumptions cannot always

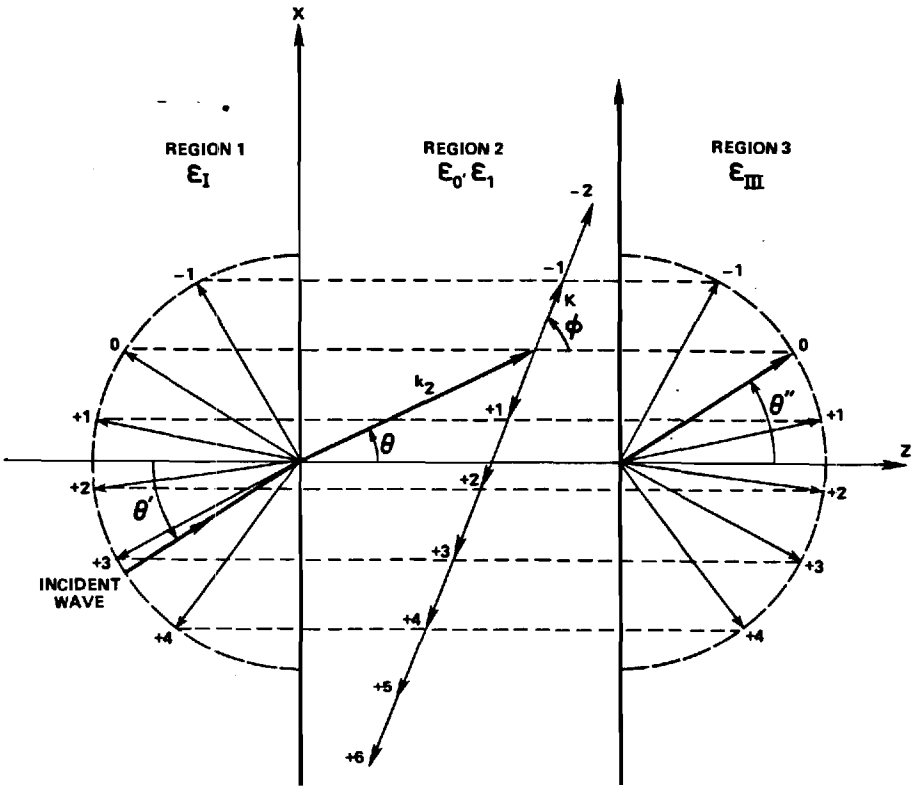


Fig. 2. Allowed wavevectors in regions 1 and 3 due to the presence of a slanted grating with wavevector $\mathbf{K}$. Phase matching of the diffracted waves outside of the grating with the boundary components of the wavevectors inside the grating (via Floquet construction) is shown. For $-1 \leqq i \leqq +4$, propagating diffracted orders exist in regions 1 and 3 whereas for $i \leqq -2$ and $i \geqq +5$, the waves are evanescent

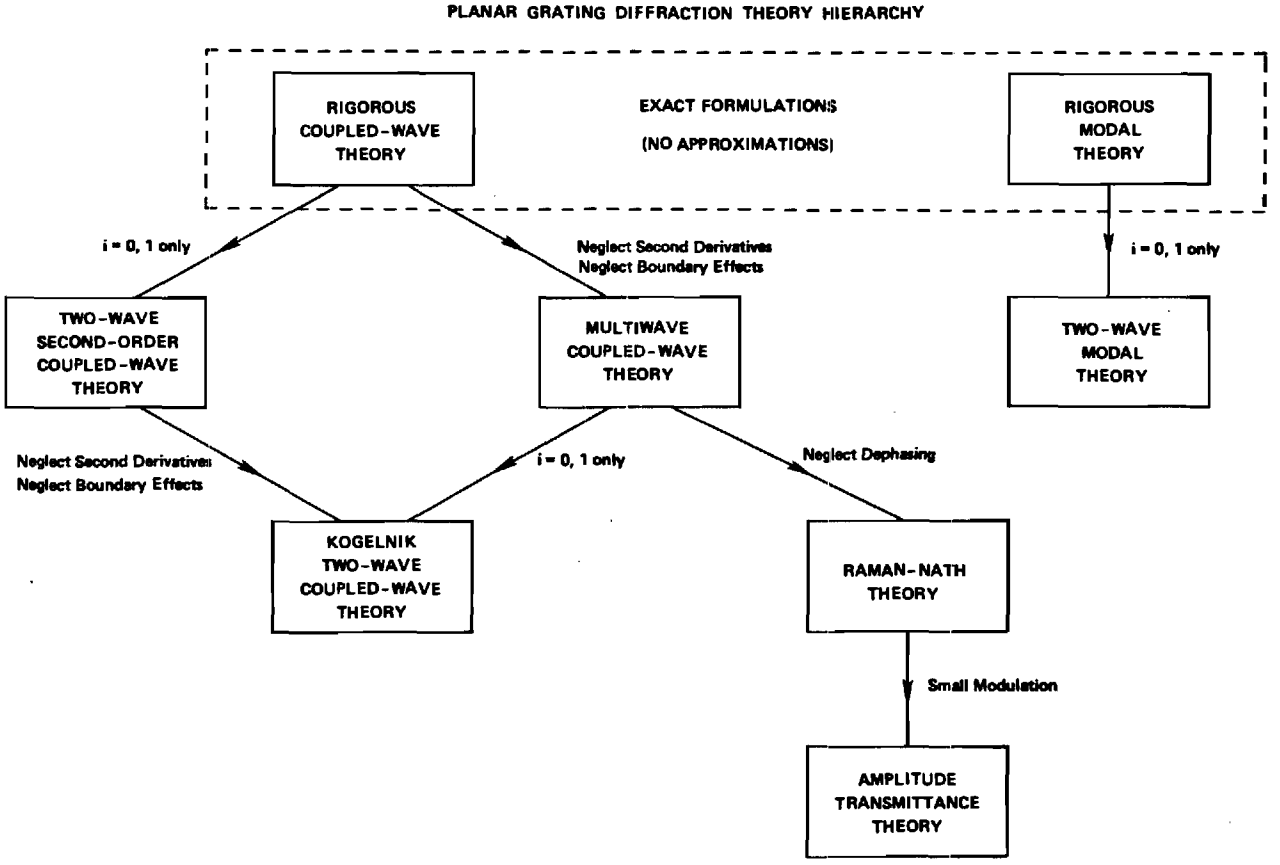


Fig. 3. Interrelationships between various planar grating diffraction theories in terms of fundamental approximations

be isolated. For example, retaining only two waves ($i=0, 1$) and the small modulation approximation can be linked for the case of a sufficiently "thick" grating.

8.1. Two-Wave Modal Theory

If only the zero and first order waves ($i=0, 1$) are retained and all higher-order waves are neglected, a two-wave regime is being assumed. There are actually a total of four waves in this analysis since there are two more waves phased matched to these. Modal theory solutions in the two-wave regime were first obtained by Bergstein and Kermisch [13] with more recent results being contributed by Lederer and Langbein [26], and Russell [19]. In this approch, the standard modal expansion (7) is used to represent the fields in the grating. However, in the two-wave case only the first two Fourier components ($i=0, 1$) of the periodic function $\Phi_m(\bar{r})$ are retained in the analysis, (25). Comparison of two-wave modal theory with exact rigorous theory [8] has shown that this can be valid near Bragg incidence in reflection gratings (backward-diffracted waves dominate). Comparison data are shown in [Ref. 8, Fig. 9].

8.2. Two-Wave Second-Order Coupled-Wave Theory

Two-wave second-order coupled-wave theory and two-wave modal theory represent exactly the same approximation. Both representations include second derivatives of field amplitudes and boundary effects. Both theories retain only the transmitted wave ($i=0$) and the fundamental diffracted wave ($i=1$) and their phased matched waves and neglect higher-order waves. This approximate theory has been used by Kong [6]. Additional approximations in this theory have been made by Kessler and Kowarschik [27–29], and by Jaaskelainen et al. [30]. The two governing equations may be obtained directly from the rigorous coupled-wave equations (11) by keeping only terms in S_0 and S_1 and neglecting all other field amplitudes. The resulting two equations from (11) are:

$$\frac{1}{2\pi^2}\frac{d^2S_0(z)}{dz^2} - j\frac{2(\varepsilon_0)^{1/2}\cos\theta}{\pi\lambda}\frac{dS_0(z)}{dz} + \frac{\varepsilon_1}{\lambda^2}S_1(z) = 0, \quad (42)$$

$$\frac{1}{2\pi^2}\frac{d^2S_1(z)}{dz^2} - j\frac{2}{\pi}\left[\frac{(\varepsilon_0)^{1/2}\cos\theta}{\lambda} - \frac{\cos\phi}{\Lambda}\right]\frac{dS_1(z)}{dz} + \frac{2(m-1)}{\Lambda^2}S_1(z) + \frac{\varepsilon_1}{\lambda^2}S_0(z) = 0. \quad (43)$$

Kong [6] has presented analytical solutions for the two-wave second-order coupled-wave theory expressed in the form of two transmission and two reflection coefficients for the unslanted-fringe planar slab grating.

8.3. Multiwave Coupled-Wave Theory

Multiwave first-order coupled-wave theory may also be developed directly from the rigorous coupled-wave equations (11). In this approach higher-order waves are retained (hence "multiwave"). The second derivatives of the field amplitudes (and thus boundary effects) are neglected. The resulting multiwave coupled-wave equations from (11) are:

$$-j\frac{2}{\pi}\left[\frac{(\varepsilon_0)^{1/2}\cos\theta}{\lambda} - \frac{i\cos\phi}{\Lambda}\right]\frac{dS_i(z)}{dz} + \frac{2i(m-i)}{\Lambda^2}S_i(z) + \frac{\varepsilon_1}{\lambda^2}[S_{i+1}(z)+S_{i-1}(z)] = 0. \quad (44)$$

For the case of an unslanted transmission grating ($\phi=\pi/2$) and normal incidence ($\theta=0$, $m=0$), the multiwave coupled-wave equations first appeared in a 1936 paper by Raman and Nath [31] for a sinusoidal (rather than cosinusoidal) grating. This paper was the fourth in a series of five papers by Raman and Nath [31–35] on the diffraction of light by sound waves. The first three papers [32–34] from the basis of the "Raman-Nath theory" described below. This simplified multiwave coupled-wave equation was referred to by Nath [36] as being due to Nath [37]. In this 1936 paper, Nath [37] obtained a very slowly converging series solution for the multi-wave coupled-wave difference-differential equations. An alternative series solution was later presented by Berry [38]. This series solution is in terms of Bessel functions and is also very slowly converging. Numerical solutions of the multiwave coupled-wave equations (also for acousto-optic interaction studies) have been obtained by Klein and Cook [3].

The multiwave coupled-wave equations have been generalized to include loss and gratings of arbitrary nonsinusoidal profile by Magnusson and Gaylord [7]. In that paper, numerical solutions were obtained for unslanted transmission gratings using a Runge-Kutta algorithm to solve the first-order system of coupled-wave equations. Diffraction efficiency results for sinusoidal, square-wave, and sawtooth phase gratings at first, second, and third Bragg incidence are presented there.

Comparison of diffraction efficiency results from multiwave first-order coupled-wave theory with exact rigorous theory has shown that this theory without second derivatives gives good results in transmission gratings (forward-diffraction waves dominate) when the grating modulation is small. Comparison data are shown in [Ref. 8, Figs. 7 and 8].

8.4. Two-Wave First-Order Coupled-Wave Theory

If higher-order waves ($i\neq 0, 1$) and second derivatives of field amplitudes (and thus boundary effects) are *both*

neglected, the rigorous coupled-wave equations(11) reduce to two-wave first-order coupled-wave theory. For general slanted gratings at arbitrary incidence the two governing equations are:

$$\cos\theta\frac{dS_0(z)}{dz}+j\frac{\pi\varepsilon_1}{2(\varepsilon_0)^{1/2}\lambda}S_1(z)=0, \tag{45}$$

$$\left(\cos\theta-\frac{\lambda}{\Lambda(\varepsilon_0)^{1/2}}\right)\frac{dS_1(z)}{dz}+j\frac{\pi\lambda(m-1)}{\Lambda^2(\varepsilon_0)^{1/2}}S_1(z)+j\frac{\pi\varepsilon_1}{2(\varepsilon_0)^{1/2}\lambda}S_0(z)=0. \tag{46}$$

Two-wave first-order coupled-wave theory was applied to acousto-optics by Phariseau [2]. It was first applied to holography by Kogelnik [4]. His thorough 1969 paper [4] is now very widely referenced. As a result, this theory is commonly called "Kogelnik theory" and this is noted in Fig. 3. The substantial recognition received by Kogelnik's paper [4] is due in part to the comprehensive coverage of 1) phase, absorption, and mixed gratings; 2) on-Bragg and off-Bragg incidence; 3) pure transmission ($\phi=\pi/2$), pure reflection ($\phi=0$), and general slanted fringe gratings; and 4) both H-mode and E-mode polarization.

From (8), if $k_2\cos\theta-K\cos\phi$ is positive, the single diffracted wave in this analysis is forward-diffracted and the grating is called a transmission grating. If $k_2\cos\theta-K\cos\phi$ is negative, the single diffracted wave is backward-diffracted and the grating is called a reflection grating. For the forward-diffracted case, the boundary condition used is $S_1(0)=0$. In the backward-diffracted case, the boundary condition used is $S_1(d)=0$. Due to the first-order nature of this theory, some phase matched waves will be neglected. In the transmission grating case, for example, the two backward-traveling waves (that are phase matched to the zero-order transmitted wave and the fundamental diffracted wave) are neglected.

For the special case of a phase grating with unslanted fringes ($\phi=\pi/2$) and incidence at the first Bragg angle ($m=1$), the first-order diffracted amplitude from (45) and (46) is given by [2, 4]

$$S_1(z)=-j\sin\left(\frac{\pi\varepsilon_1 z}{2(\varepsilon_0)^{1/2}\lambda\cos\theta}\right), \tag{47}$$

where z is the distance into the grating at which the amplitude is determined. This well-known expression predicts a diffraction efficiency [$\mathrm{DE}=S_i(d)S_i^*(d)$ for this case] that is sinusoidal in modulation and has a maximum value of 100%. Although the two-wave first-order coupled-wave theory neglects higher-order diffracted waves and second derivatives of field amplitudes (and thus also boundary effects), it nevertheless contains many of the basic features of the diffraction process in an extended grating. This theory has been sucessfully extended to numerous other cases including finite beams [39, 40], finite and nonplanar gratings [41–43], and attenuated gratings [29, 44–46]. When grating diffraction is described by the two-wave result, (47), it is often referred to as "Bragg regime" diffraction. Incidence at the Bragg angle is essential in "Bragg regime" diffraction whereas in "Raman Nath regime" diffraction described below it is not. Criteria for "Bragg regime" behavior are given in [47].

For the basic case of a uniform grating and plane wave, a comparison of diffraction efficiency results from two-wave first-order coupled-wave theory with exact rigorous theory is presented in [8] for a series of grating slant angles. When transmission grating behavior dominates, the error in two-wave coupled-wave theory. is due primarily to the neglect of higher-order waves in the theory. Conversely, when reflection grating behavior dominates, the error is primarily due to the neglect of second derivatives and boundary effects.

8.5. Raman-Nath Theory

The theory of Raman and Nath [32–35] may also be obtained directly from the rigorous coupled-wave equations. If second derivatives of the field amplitudes and dephasing from the Bragg condition are both neglected, the rigorous coupled-wave equations (11) reduce to the Raman-Nath diffraction equations:

$$-j\frac{2}{\pi}\left[\frac{(\varepsilon_0)^{1/2}\cos\theta}{\lambda}-\frac{i\cos\phi}{\Lambda}\right]\frac{dS_i(z)}{dz}+\frac{\varepsilon_1}{\lambda^2}[S_{i+1}(z)+S_{i-1}(z)]=0, \tag{48}$$

where a general angle of incidence and grating slant angle have been retained. The S_i term in (11) has been neglected. For the i-th diffracted order, this term is zero for the m-th Bragg incidence, (12), when $i=m$. For an arbitrary angle of incidence, each diffracted order will be dephased by varying amounts from their corresponding Bragg conditions. This in turn reduces the synchronism between the input wave and that diffracted order. The result is less coupling from the input to that order. The Raman-Nath theory therefore, treats all diffracted orders as though the Bragg conditions for all them were simultaneously satisfield.

For the important case of an unslanted fringe transmission grating ($\phi=\pi/2$), (48) takes the form of a recurrence relation satisfied by Bessel functions. The solution is

$$S_i(z)=(-j)^i J_i\left(\frac{\pi\varepsilon_1 z}{(\varepsilon_0)^{1/2}\lambda\cos\theta}\right) \tag{49}$$

for boundary conditions $S_0(0)=1$ and $S_i(0)=0$ ($i\neq 0$) where J_i is an integer-order Bessel function of the

first kind. Equation (49) is the famous Bessel function expression of Raman and Nath. It predicts maximum diffraction efficiencies of $DE_{\pm 1}=33.8\%$, $DE_{\pm 2}=23.6\%$, $DE_{\pm 3}=18.8\%$, and so forth. When grating diffraction behavior may be approximated by (49), it is referred to as "Raman-Nath regime" diffraction. This result, (49), has been extensively used to predict the light intensities diffracted by sound waves [38, 48]. Criteria for "Raman-Nath regime" diffraction are given in [49]. Raman-Nath theory has been extended to describe nonsinusoidal phase gratings [50–52].

8.6. *Amplitude Transmittance Theory*

For gratings, the amplitude transmittance approach is closely related to Raman-Nath diffraction theory. The amplitude transmittance approach is widely used in optics [53–54] and may be applied to slabs, lenses, apertures, and general two-dimensional objects as well as gratings. The amplitude transmittance is defined as the ratio of the field amplitude over the output plane to the field amplitude incident on the input plane. The amplitude transmittance function in general is complex. It may be applied to gratings with unslanted fringes. Both amplitude gratings [53–57] and phase gratings [51–56] have been treated in the literature using the amplitude transmittance approach.

For a phase grating with periodicity in the x direction, the amplitude transmittance function is

$$\tau(x,z)=\exp\left(-j\frac{2\pi[\varepsilon(x)]^{1/2}z}{\lambda\cos\theta}\right), \tag{50}$$

where z is the grating thickness and $n(x)=[\varepsilon(x)]^{1/2}$ is the periodic refractive index. Since the transmittance function is also periodic in x, it may be expanded in a complex Fourier series. Further, because the exponentials in this series are in the form of an expansion of the diffracted plane waves, then the Fourier coefficients are the diffracted wave amplitudes. The Fourier series expansion may thus be written

$$\tau(x,z)=\sum_i S_i(z)\exp(\mathrm{j}iKx), \tag{51}$$

where S_i represents the amplitude of the i-th diffracted order. By definition, the coefficients of the Fourier series may be calculated from

$$S_i(z)=\frac{1}{\Lambda}\int_0^{\Lambda}\exp\left(-\mathrm{j}\frac{2\pi[\varepsilon(x)]^{1/2}z}{\lambda\cos\theta}\right)\exp(-\mathrm{j}iKx)dx. \tag{52}$$

Thus the diffracted amplitudes may be determined directly knowing $\varepsilon(x)$ by integrating (52). Results for sinusoidal, square-wave, sawtooth, triangular, and rectangular refractive-index profiles are given in [51].

For the unslanted-fringe (co)sinusoidal-permittivity transmission grating, the corresponding index of refraction is

$$n(x)\equiv[\varepsilon(x)]^{1/2}=(\varepsilon_0+\varepsilon_1\cos Kx)^{1/2} \tag{53}$$

which may be expanded in a Fourier cosine series as

$$[\varepsilon(x)]^{1/2}=[\varepsilon(x)]_0^{1/2}+\sum_{h=1}^{\infty}[\varepsilon(x)]_h^{1/2}\cos(hKx) \tag{54}$$

with Fourier harmonic amplitudes given by

$$[\varepsilon(x)]_h^{1/2}=\frac{2}{\Lambda}\int_0^{\Lambda}(\varepsilon_0+\varepsilon_1\cos Kx)^{1/2}\cos(hKx)dx. \tag{55}$$

The average value of the refractive index may be expressed concisely as

$$n_0(x)\equiv[\varepsilon(x)]_0^{1/2}=(2/\pi)(\varepsilon_0+\varepsilon_1)^{1/2}E(\zeta,\pi/2), \tag{56}$$

where $E(\zeta,\pi/2)$ is the complete elliptic integral of the second kind and $\zeta\equiv 2\varepsilon_1/(\varepsilon_0+\varepsilon_1)$. Clearly, the case of a sinusoidal permittivity (or dielectric constant) being treated throughout this paper is not the same as a sinusoidal refractive-index grating. The index of refraction corresponding to sinusoidal permittivity has higher spatial frequency harmonics ($h>1$) in addition to a fundamental sinusoidal component ($h=1$) as represented by (54). However, for the case of sufficiently small modulation, a sinusoidal permittivity produces nearly a sinusoidal index of refraction. In the limit of small modulation (ε_1 approaches zero), (55) and (56) yield

$$[\varepsilon(x)]_0^{1/2}\simeq\varepsilon_0, \tag{57}$$

$$[\varepsilon(x)]_1^{1/2}\simeq\varepsilon_1/2(\varepsilon_0)^{1/2}, \tag{58}$$

$$[\varepsilon(x)]_2^{1/2},\quad [\varepsilon(x)]_3^{1/2},\ldots\simeq 0. \tag{59}$$

This analysis is important in that it now allows the Raman-Nath theory and amplitude transmittance theory to be interrelated. The result is that although (52) was obtained using the amplitude transmittance approach, it is *also* a solution of the Raman-Nath difference-differential equation (48) for unslanted gratings in the limit of small modulation. This may be shown by direct substitution of S_i as given by (52) into the Raman-Nath diffraction equation (48). Thus, for a cosinusoidal refractive-index profile, the integral in (52) when evaluated gives the Bessel function result (49). This may be accomplished using the identity

$$\exp(-\mathrm{j}b\cos\alpha)\equiv\sum_{i=-\infty}^{+\infty}(-\mathrm{j})^iJ_i(b)\exp(\mathrm{j}i\alpha) \tag{60}$$

and the orthogonality relationship

$$\frac{1}{\Lambda}\int_0^{\Lambda}\exp(\mathrm{j}lKx)\exp(-\mathrm{j}iKx)dx=\delta_{li}, \tag{61}$$

where δ_{ii} is the Kronecker delta. Therefore, as depicted in Fig. 3, it has been shown that Raman-Nath theory and amplitude transmittance theory are equivalent in the limit of small grating modulation. This is true in general for unslanted gratings regardless of the grating profile (squarewave, sawtooth, etc.).

9. Discussion and Conclusions

Theories describing the diffraction of a plane electromagnetic wave incident with arbitrary wavelength and angle of incidence upon a slanted fringe planar sinusoidal permittivity grating have been reviewed. For simplicity the analysis has been restricted to H-mode polarization (electric field perpendicular to plane of incidence). Exact formulations without approximations (rigorous coupled-wave analysis and rigorous modal analysis) have been developed and shown to be mathematically equivalent. Rigorous coupled-wave equations have been developed in alternative forms, (11) and (14), and their usefulness discussed. The solution of the rigorous coupled-wave equations (11) in terms of state variables has been presented in detail along with the phase matching and boundary conditions necessary to determine the diffracted amplitudes outside of the grating.

These rigorous theories have been shown to reduce to the approximate theories: 1) two-wave modal theory, 2) two-wave second-order coupled-wave theory, 3) multiwave first-order coupled-wave theory, 4) two-wave first-order coupled-wave theory (Kogelnik theory), 5) Raman-Nath theory, and 6) amplitude transmittance theory in the appropriate limits. The assumptions associated with each of these approximate theories have been explicitly presented.

The rigorous theories presented in this paper (and thus their approximate versions) are based on the Floquet theorem. As such, they require a truly periodic grating (an infinite number of periods). These theories may be applied in the angular limit as the slanted fringes of a general grating approach being parallel to the surface (ϕ approaches zero). However, for exact parallelism with the surface ($\phi = 0$), the grating is no longer strictly periodic and a continuum of solutions is possible depending on the number of periods, the starting conditions, and the ending conditions of the grating. This pure reflection grating case can be analyzed without approximation using a rigorous chain-matrix method of analysis. This is discussed in [58].

Acknowledgements. This work was sponsored by the National Science Foundation and by the Joint Services Electronics Program.

References

1. R.R. Aggarwal: Proc. Indian Acad. Sci. **31**, 417–426 (1950)
2. P. Phariseau: Proc. Indiana Acad. Sci. **A44**, 165–170 (1956)
3. W.R. Klein, B.D. Cook: IEEE Trans. SU**14**, 123–134 (1967)
4. H. Kogelnik: Bell Syst. Tech. J. **48**, 2909–2947 (1969)
5. G.L. Fillmore, R.F. Tynan: J. Opt. Soc. Am. **61**, 199–203 (1971)
6. J.A. Kong: J. Opt. Soc. Am. **67**, 825–829 (1977)
7. R. Magnusson, T.K. Gaylord: J. Opt. Soc. Am. **67**, 1165–1170 (1977)
8. M.G. Moharam, T.K. Gaylord: J. Opt. Soc. Am. **71**, 811–818 (1981)
9. T. Tamir, H.C. Wang, A.A. Oliner: IEEE Trans. MTT-**12**, 323–335 (1964)
10. T. Tamir, H.C. Wang: Can. J. Phys. **44**, 2073–2094 (1966)
11. T. Tamir: Can. J. Phys. **44**, 2461–2494 (1966)
12. C.B. Burckhardt: J. Opt. Soc. Am. **56**, 1502–1509 (1966)
13. L. Bergstein, D. Kermisch: Proc. Symp. Mod. Opt. **17**, 655–680 (1967)
14. R.S. Chu, T. Tamir: IEEE Trans. MTT-**18**, 486–504 (1970)
15. R.S. Chu, T. Tamir: Proc. IEE **119**, 797–806 (1972)
16. F.G. Kaspar: J. Opt. Soc. Am. **63**, 37–45 (1973)
17. S.T. Peng, T. Tamir, H. L. Bertoni: IEEE Trans. MTT-**23**, 123–133 (1975)
18. R.S. Chu, J.A. Kong: IEEE Trans. MTT-**25**, 18–24 (1977)
19. P.St.J. Russell: Phys. Rep. **71**, 209–312 (1981)
20. L.Solymar, D.J. Cooke: *Volume Holography and Volume Gratings* (Academic Press, London 1981)
21. R. Magnusson, T.K. Gaylord: J. Opt. Soc. Am. **68**, 1777–1779 (1978)
22. D.A. Watkins: *Topics in Electromagnetic Theory* (Wiley, New York 1958) p. 2
23. R.E. Collin: *Field Theory of Guided Waves* (McGraw-Hill, New York 1960) p. 368
24. E.g.C.L. Liu, J.W.S. Liu: *Linear Systems Analysis* (McGraw-Hill, New York 1975)
25. E.g. program EIGRF from the International Mathematics and Statistics Library (IMSL)
26. F. Lederer, U. Langbein: Opt. Acta **27**, 183–200 (1980)
27. S. Kessler, R. Kowarschik: Opt. Quantum Electron. **7**, 1–4 (1975)
28. R. Kowarschik, S. Kessler: Opt. Quantum Electron. **7**, 399–411 (1975)
29. R. Kowarschik: Opt. Acta **23**, 1039–1051 (1976)
30. T. Jaaskelainen, O. Salminen, R.M.K. Hamalainen: Acta Poly.-Appl. Phys. **126**, 3–12 (1979)
31. C.V. Raman, N.S.N. Nath: Proc. Indian Acad. Sci. **A3**, 119–125 (1936)
32. C.V. Raman, N.S.N. Nath: Proc. Indian Acad. Sci. **A2**, 406–412 (1935)
33. C.V. Raman, N.S.N. Nath: Proc. Indian Acad. Sci. **A2**, 413–420 (1935)
34. C.V. Raman, N.S.N. Nath: Proc. Indian Acad. Sci. **A3**, 75–84 (1936)
35. C.V. Raman, N.S.N. Nath: Proc. Indian Acad. Sci. **A3**, 459–465 (1936)
36. N.S.N. Nath: Proc. Indian Acad. Sci. **A8**, 499–503 (1938)
37. N.S.N. Nath: Proc. Indian Acad. Sci. **A4**, 222–242 (1936)
38. M.V. Berry: *The Diffraction of Light by Ultrasound* (Academic Press, London 1966)
39. P.St.J. Russell, L. Solymar, M.P. Jordan: Proc. ICO **11**, 635–638 (1978)
40. M.G. Moharam, T.K. Gaylord, R. Magnusson: J. Opt. Soc. Am. **70**, 300–304 (1980)
41. L. Solymar, M.P. Jordan: Microwave Opt. Acoust. **1**, 89–92 (1977)
42. R.P. Kenan: IEEE J. QE-**14**, 924–930 (1978)
43. M.G. Moharam, T.K. Gaylord, R. Magnusson: J. Opt. Soc. Am. **70**, 437–442 (1980)
44. N. Uchida: J. Opt. Soc. Am. **63**, 280–287 (1973)
45. U. Killat: Opt. Commun. **21**, 110–111 (1977)
46. M.P. Owen, L. Solymar: Opt. Commun. **34**, 321–326 (1980)

47. M.G. Moharam, T.K. Gaylord, R. Magnusson: Opt. Commun. **32**, 19–23 (1980)
48. F.H. Sanders: Can. J. Phys. **14**, 158–171 (1936)
49. M.G. Moharam, T.K. Gaylord, R. Magnusson: Opt. Commun. **32**, 14–18 (1980)
50. K.L. Zankel, E.A. Hiedemann: J. Acoust. Soc. Am. **31**, 44–54 (1959)
51. R. Magnusson, T.K. Gaylord: J. Opt. Soc. Am. **68**, 806–809 (1978)
52. R. Magnusson, T.K. Gaylord: Opt. Commun. **25**, 129–132 (1978).
53. J.W. Goodman: *Introduction to Fourier Optics* (McGraw-Hill, New York 1968)
54. W.T. Cathey: *Optical Information Processing and Holography* (Wiley, New York 1974)
55. H.M. Smith: *Principles of Holography* (Wiley, New York 1969)
56. R.J. Collier, C.B. Burckhardt, L.H. Lin: *Optical Holography* (Academic Press, New York 1971)
57. R. Magnusson, T.K. Gaylord: Opt. Commun. **28**, 1–3 (1979)
58. M.G. Moharam, T.K. Gaylord: J. Opt. Soc. Am. **72**, 187–190 (1982)

Diffraction analysis of dielectric surface-relief gratings

M. G. Moharam and T. K. Gaylord

School of Electrical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

Received January 11, 1982; revised manuscript received May 3, 1982

Diffraction by a dielectric surface-relief grating is analyzed using rigorous coupled-wave theory. The analysis applies to arbitrary grating profiles, groove depths, angles of incidence, and wavelengths. Example results for a wide range of groove depths are presented for sinusoidal, square-wave, triangular, and sawtooth gratings. Diffraction efficiencies obtained from the present method of analysis are compared with previously published numerical results. To obtain large diffraction efficiencies (greater than 85%) for gratings with typical substrate permittivities, it is shown that the grating profile should possess even symmetry.

1. INTRODUCTION

Dielectric surface-relief (or corrugated) gratings are of wide interest owing to their many applications in quantum electronics, integrated optics, spectroscopy, and holography.[1] Example devices include distributed-feedback lasers, distributed-Bragg-reflector lasers, beam deflectors, waveguide couplers, spectral filters, wavelength multiplexers and demultiplexers, holographic beam combiners, and others. Several investigators have obtained rigorous solutions for the exact electromagnetic boundary-value problem that apply to gratings with rectangular or triangular grooves.[2-8] Others have obtained approximate results by applying perturbation techniques.[9,10] For gratings with arbitrary profiles, the integral method[1,11-13] was first used to obtain numerical results. Later the differential method was developed,[1,2] and it is probably the most widely used method for analyzing arbitrary-profile gratings.[2,14-16] The differential method requires advanced numerical techniques, such as the combined Runge–Kutta and Adam–Moulton algorithm,[1,2] the unimoment method,[14] Numerov's algorithm,[1] the modified Adam–Moulton algorithm,[16] and other similar algorithms. Some of these algorithms do not converge for relatively deep grooves.

Recently, a rigorous coupled-wave approach was formulated and applied to planar dielectric gratings.[17] A state-variables representation of the coupled-wave amplitudes permits these amplitudes to be obtained in terms of the eigenfunctions and eigenvectors of the coefficient matrix defined by the rigorous coupled-wave equations.

In this paper, the rigorous coupled-wave approach is adapted to the exact electromagnetic boundary-value problem associated with dielectric surface-relief gratings of arbitrary profile. In this method, a surface-relief grating is divided into a large number of thin layers parallel to the surface. Each thin grating is then analyzed by using the state-variables method of solution of the rigorous coupled-wave equations for that grating. By formulating the problem in a particular manner, it is shown that the grating layers may be treated one at a time in sequence, thus reducing the numerical calculations to an easily manageable size. There are no approximations in the analysis, and results are obtainable to an arbitrary level of accuracy. The diffraction efficiencies of all orders of both the transmitted and reflected waves are determined in the process. Peng *et al.*[5] were the first to propose partitioning the grating into thin layers and approximating each layer by a rectangular profile. They then calculated the field in each layer by using the Lewis and Hessel[18] method of analyzing a periodic array of slabs. However, they presented numerical results only for one-[3] and two-layered[4] gratings. In the present work, the rigorous coupled-wave approach is used to find the fields in each layer, and a large number of layers (typically 50–100) are used.

To illustrate the utility of this method of analysis, the diffraction characteristics of sinusoidal, square-wave, triangular, and sawtooth gratings are determined for a wide range of groove depths for a plane wave incident at the first Bragg angle. Very large groove depths (4 grating periods deep for the examples presented) may be handled routinely with this analysis without any type of numerical difficulty. In a final section the accuracy and the convergence of the present method are discussed in detail. The diffraction-efficiency results obtained by the present method are compared with numerical results obtained by other techniques.[12-14,16] It is shown that the rigorous coupled-wave method accurately predicts conservation of power when all the individual diffracted powers are summed.

2. THEORY

The general dielectric surface-relief grating diffraction problem treated in this paper is shown in Fig. 1. An electromagnetic plane wave is obliquely incident upon the grating. The periodic structure produces, in general, both forward-diffracted and backward-diffracted waves as shown in the figure. Region 1 (the input region) is a homogeneous dielectric with a relative permittivity (dielectric constant) of ϵ_I. Likewise, region 3 is homogeneous with a relative permittivity of ϵ_{III}. Region 2 (the grating region) consists of a periodic distribution of both types of dielectrics. The boundary between the ϵ_I dielectric and the ϵ_{III} dielectric in region 2 is given by

$$z = F(x) = F(x + \Lambda), \tag{1}$$

where Λ is the grating period. The function $F(x)$ thus represents the grating-surface profile. There are no restrictions

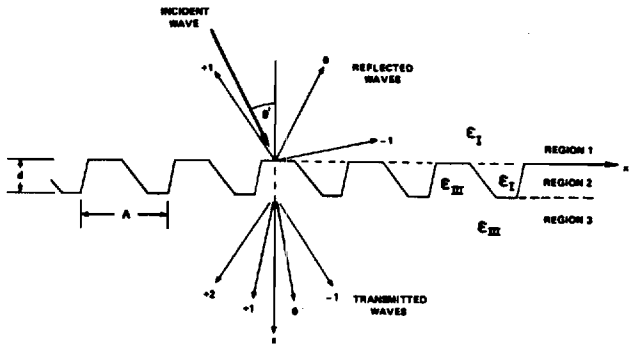


Fig. 1. Geometry of dielectric surface-relief grating.

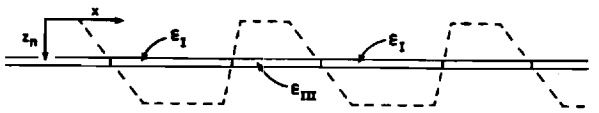


Fig. 2. The nth planar grating resulting from the decomposition of the surface-relief grating into N thin gratings.

on the form of $F(x)$ in this analysis. Curved lines, straight lines, shadow regions, etc. are all allowed. In this paper, for simplicity, the case of lossless dielectrics and incident plane-wave polarization perpendicular to the plane of incidence (H mode) are treated.

The total electric field in region 1 is the sum of the incident and the backward-traveling waves. The normalized total electric field in region 1 may be expressed as

$$E_1 = \exp(-j\mathbf{k}_1 \cdot \mathbf{r}) + \sum_{i=-\infty}^{\infty} R_i \exp(-j\mathbf{k}_{1i} \cdot \mathbf{r}), \tag{2}$$

where $j = (-1)^{1/2}$, $\mathbf{k}_1$ is the incident-field wave vector of magnitude, $k_1 = 2\pi(\epsilon_I)^{1/2}/\lambda$, λ is the free-space wavelength, and $\mathbf{r} = x\hat{x} + y\hat{y} + z\hat{z}$, with $\hat{}$ denoting a unit vector. R_i is the normalized amplitude of the ith reflected wave in region 1 with wave vector $\mathbf{k}_{1i}$. Likewise, the normalized total electric field in region 3 is

$$E_3 = \sum_{i=-\infty}^{\infty} T_i \exp[-j\mathbf{k}_{3i} \cdot (\mathbf{r} - d\hat{z})], \tag{3}$$

where T_i is the normalized amplitude of the ith transmitted wave into region 3 with wave vector $\mathbf{k}_{3i}$ and d is the groove depth. The quantities $\mathbf{k}_{1i}$ and $\mathbf{k}_{3i}$ are determined below by using the phase-matching requirement.

In the present analysis, the grating region (region 2) is divided into N thin planar-grating slabs perpendicular to the z axis. Then the rigorous coupled-wave analysis that has been developed for planar gratings is applied to each slab grating.[17] If the individual planar gratings are sufficiently thin, any grating profile can be analyzed to an arbitrary level of accuracy. The nth slab within region 2 as shown in Fig. 2 will consist of a periodic distribution of ϵ_I and ϵ_{III} dielectrics. The relative permittivity for the nth slab grating is periodic, $\epsilon_n(x, z_n) = \epsilon_n(x + \Lambda, z_n)$, and may be expanded in a Fourier series as

$$\epsilon_n(x, z_n) = \epsilon_I + (\epsilon_{III} - \epsilon_I) \sum_{h=-\infty}^{+\infty} \tilde{\epsilon}_{h,n} \exp(jhKx), \tag{4}$$

where z_n is the z coordinate of the nth slab, h is the harmonic index, K is the magnitude of the grating vector ($K = 2\pi/\Lambda$), and $\tilde{\epsilon}_{h,n}$ are the normalized complex harmonic amplitude coefficients given by

$$\tilde{\epsilon}_{h,n} = (1/\Lambda) \int_0^{\Lambda} f(x, z_n) \exp(-jhKx)dx, \tag{5}$$

where the function $f(x, z_n)$ is equal to either zero or unity, depending on whether, for a particular value of x, the grating relative permittivity is ϵ_I or ϵ_{III}, respectively.

In the coupled-wave representation used in this treatment, the fields inside each grating slab are expanded in terms of the space harmonics of the fields in the periodic structure. These space harmonics inside the grating correspond to diffracted orders outside the grating. Thus the partial fields inside the modulated medium are visualized as diffracted waves that progress through the planar slab and couple energy back and forth between each other as they progress. In the coupled-wave approach the total field is thus expressed as

$$E_{2,n} = \sum_{i=-\infty}^{+\infty} S_{i,n}(z) \exp(-j\sigma_{i,n} \cdot \mathbf{r}), \tag{6}$$

where i is the space-harmonic index (integer) and $S_{i,n}(z)$ are the space-harmonic field amplitudes. As a result of the Floquet theorem, the diffracted wave vectors inside the nth grating, $\sigma_{i,n}$, may be represented by

$$\sigma_{i,n} = \mathbf{k}_{2,n} - i\mathbf{K}, \tag{7}$$

where $\mathbf{K} = K\hat{x}$ and $\mathbf{k}_{2,n}$ is the wave vector of the zero-order ($i = 0$) refracted wave having a magnitude of $k_{2,n} = 2\pi(\epsilon_{0,n})^{1/2}/\lambda$, and $\epsilon_{0,n}$ is the average relative permittivity for the nth slab grating.

Each ith diffracted field in regions 1 and 3 must be phase matched to the ith space-harmonic field inside each nth slab grating. Thus $\mathbf{k}_{1i} \cdot \hat{x} = (\mathbf{k}_{2,n} - i\mathbf{K}) \cdot \hat{x} = \mathbf{k}_{3i} \cdot \hat{x}$ for any i and n. In the homogeneous regions (1 and 3) the backward- and forward-diffracted waves have wave vectors with magnitudes $|\mathbf{k}_{1i}| = |\mathbf{k}_1|$ and $|\mathbf{k}_{3i}| = |\mathbf{k}_3|$, where $k_3 = 2\pi(\epsilon_{III})^{1/2}/\lambda$. Knowing the total amplitudes and the x components of the diffracted wave vectors, the z components are then determined to be $\mathbf{k}_{1i} \cdot \hat{z} = [|\mathbf{k}_1|^2 - (\mathbf{k}_{1i} \cdot \hat{x})^2]^{1/2}$ and $\mathbf{k}_{3i} \cdot \hat{z} = [|\mathbf{k}_3|^2 - (\mathbf{k}_{3i} \cdot \hat{x})^2]^{1/2}$. Therefore the total fields in all regions may be rewritten as

$$E_1 = \exp\{-j[k_1(\sin\theta' x + \cos\theta' z)]\} + \sum_{i=-\infty}^{+\infty} R_i \exp(-j\{(k_1 \sin\theta' - iK)x - [k_1^2 - (k_1 \sin\theta' - iK)^2]^{1/2} z\}), \tag{8}$$

$$E_{2,n} = \sum_{i=-\infty}^{+\infty} S_{i,n}(z) \exp\{-j[(k_1 \sin\theta' - iK)x + (k_{2,n}^2 - k_1^2 \sin^2\theta')^{1/2} z]\}, \tag{9}$$

and

$$E_3 = \sum_{i=-\infty}^{\infty} T_i \exp(-j\{(k_1 \sin\theta' - iK)x + [k_3^2 - (k_1 \sin\theta' - iK)^2]^{1/2}(z - d)\}). \tag{10}$$

In region 2, the wave equation for H-mode polarization

$$\nabla^2 E_{2,n} + k^2 \epsilon_n(x, z_n) E_{2,n} = 0, \tag{11}$$

where $k = 2\pi/\lambda$, must be satisfied for each nth slab grating. Substituting $\epsilon_n(x, z_n)$ from Eq. (4) and $E_{2,n}$ from Eq. (9) into the wave equation, performing the indicated differentiations,

and setting the coefficient of each exponential term equal to zero for nontrivial solutions yields the rigorous coupled-wave equations for the nth slab grating:

$$\frac{d^2S_{i,n}(z)}{dz^2} - j2(k_{2,n}^2 - k_1^2 \sin^2\theta')^{1/2}\frac{dS_{i,n}(z)}{dz} + K^2 i(m-i)S_{i,n}(z) + k^2(\epsilon_{III} - \epsilon_I)\sum_{h=1}^{\infty}[\tilde{\epsilon}_{h,n}S_{i-h,n}(z) + \tilde{\epsilon}_{h,n}^{*}S_{i+h,n}(z)] = 0. \quad (12)$$

These coupled-wave equations are an infinite set of second-order coupled difference-differential equations. Each diffracted wave (i) is coupled to other diffracted waves through the harmonics of the grating ($i-h$ and $i+h$). The quantity m has been defined as

$$m = 2\Lambda(\epsilon_I)^{1/2}\sin\theta'/\lambda. \quad (13)$$

For the case when m is an integer, this represents a Bragg condition.

3. METHOD OF SOLUTION

The dielectric surface-relief grating diffraction problem as formulated in Section 2 will be solved in a sequence of steps. First, the rigorous coupled-wave equations will be solved for the nth slab grating by using a state-variables method of solution. Second, electromagnetic boundary conditions (continuity of tangential E and tangential H) will be applied between region 1 and the first slab grating, then between the first and second slab gratings, and so forth, and finally between the Nth slab grating and region 3. Third, the resulting array of boundary-condition equations is solved for the reflected and transmitted diffracted amplitudes, R_i and T_i. From these amplitudes, the diffracted efficiencies are determined directly.

By using the methods of linear systems analysis,[19] the coupled-wave differential-equation description of the diffraction problem may be transformed into a state-space description. Defining the state variables for the nth slab grating as

$$S_{1,i,n}(z) = S_{i,n}(z), \quad (14)$$

$$S_{2,i,n}(z) = dS_{i,n}(z)/dz \quad (15)$$

transforms the infinite set of second-order differential equations [Eq. (12)] into two infinite sets of first-order state equations:

$$\frac{dS_{1,i,n}(z)}{dz} = S_{2,i,n}(z), \quad (16)$$

$$\frac{dS_{2,i,n}(z)}{dz} = -k^2(\epsilon_{III} - \epsilon_I)\sum_{h=1}^{\infty}\tilde{\epsilon}_{h,n}S_{1,i-h,n}(z) - K^2 i(m-i)S_{1,i,n}(z) - k^2(\epsilon_{III} - \epsilon_I)\sum_{h=1}^{\infty}\tilde{\epsilon}_{h,n}^{*}S_{1,i+h,n}(z) + j2(k_{2,n}^2 - k_1^2\sin^2\theta')S_{2,i,n}(z). \quad (17)$$

In matrix form, the state equation for the nth slab grating may be written as

$$\begin{bmatrix} \vdots \\ \dot{S}_{1,p,n} \\ \vdots \\ \dot{S}_{2,p,n} \\ \vdots \end{bmatrix} = \left[\begin{array}{c|c} & \\ a_{p,q,n} & b_{p,q,n} \\ & \\ \hline & \\ c_{p,q,n} & d_{p,q,n} \\ & \end{array}\right] \begin{bmatrix} \vdots \\ S_{1,q,n} \\ \vdots \\ S_{2,q,n} \\ \vdots \end{bmatrix}, \quad (18)$$

where $S_{l,p,n} \equiv S_{l,i,n}$ (for $l = 1, 2$), $\dot{S} = dS/dz$, and the elements of the four submatrices ($p = 1$ to s and $q = 1$ to s) are specified by Eqs. (16) and (17) for the nth slab grating. The integers p and q are the row and column indices of the four submatrices. The maximum value of these indices, s, is equal to the number of diffracted orders retained in the analysis. The value $p = 1$ corresponds to the most negative order (value of i) retained in the analysis, and $p = s$ corresponds to the most positive order retained. For example, if an odd number of waves are retained symmetrically about $i = 0$ (the undiffracted wave) in the analysis, then $p = i + (s+1)/2$. Equation (18) corresponds to an unforced-state equation and may be expressed concisely as $\dot{\mathbf{S}} = \mathbf{AS}$, where $\dot{\mathbf{S}}$ and $\mathbf{S}$ are the column vectors from Eq. (14) and $\mathbf{A}$ is the total coefficient matrix. Although $\mathbf{A}$ is an infinite matrix, results may be obtained in practice to an arbitrary level of accuracy with a truncated matrix. Each of the four submatrices is truncated to $s \times s$. As the integer s increases, the calculated results rapidly converge to the exact results. The solutions of Eq. (18) are[19]

$$S_{p',n}(z) = \sum_{q'=1}^{2s} C_{q',n}w_{p',q',n}\exp(\lambda_{q',n}z), \quad (19)$$

where $S_{l,p,n}$ (for $l = 1, 2$) has been rewritten as $S_{p',n}$, with $p' = p + (l-1)s$. The quantities $\lambda_{q',n}$ and $w_{p',q',n}$ are the eigenvalues and eigenvectors of the matrix $\mathbf{A}$. These values are typically calculated by using a computer library program.[20] The integers p' and q' are the row and column indices of the eigenvector matrix $[w]$ and $p' = 1$ to $2s$ and $q' = 1$ to $2s$. The quantities $C_{q',n}$ are unknown constants to be determined by the boundary conditions. The desired diffracted-wave amplitudes for the nth grating layer are given by

$$S_{i,n}(z) = S_{p',n}(z), \quad (20)$$

where p' is chosen to correspond to the ith diffracted wave.

Electromagnetic boundary conditions require that the tangential electric and tangential magnetic fields be continuous across the boundaries between the slabs. For the H-mode polarization described in this paper, the electric field only has a tangential component (y direction). The tangential component of H is in the x direction, and from Maxwell's equations it is given by $H_x = (-j/\omega\mu)E_y/\partial z$. Therefore, for the boundary ($z = 0$) between region 1 (the input region) and the first slab grating, the boundary condition for tangential E is

$$\delta_{i0} + R_i = \sum_{q'=1}^{2s} C_{q',1}w_{p',q',1}, \quad (21)$$

and the boundary condition for tangential H is

$$j(\mathbf{k}_{1i}\cdot\hat{z})(R_i - \delta_{i0}) = \sum_{q'=1}^{2s} C_{q',1}w_{p',q',1}[\lambda_{q',1} - j(\sigma_{i,1}\cdot\hat{z})], \quad (22)$$

where δ_{i0} is the Kronecker delta function and the value of p'

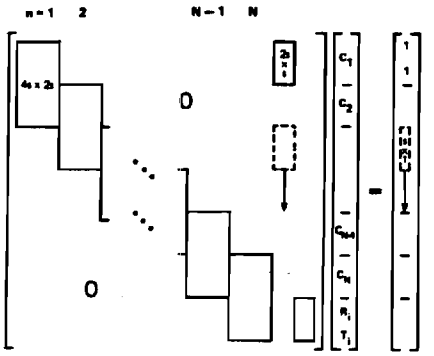


Fig. 3. Matrix-equation representation of $2(N+1)s$ boundary-condition equations, where s is the total number of diffracted waves retained in the analysis. C_1 represents the column vector $C_{q',1}$, where $q' = 1$ to $2s$ and likewise for C_2 through C_N. The reflected and transmitted amplitudes R_i and T_i are column vectors of length s. The product output vector, before manipulation, is all zeros except for the two ones that are shown. These correspond to the normalized E and H values in the input wave.

is chosen to correspond to the ith wave. For the boundary between the nth and $n+1$th slab gratings ($z = nd/N$), the boundary condition for tangential E is

$$\sum_{q'=1}^{2s} C_{q',n} w_{p',q',n} \exp\{[\lambda_{q',n} - j(\sigma_{i,n}\cdot\hat{z})]nd/N\} = \sum_{q'=1}^{2s} C_{q',n+1} w_{p',q',n+1} \exp\{[\lambda_{q',n+1} - j(\sigma_{i,n+1}\cdot\hat{z})]nd/N, \tag{23}$$

and the boundary condition for tangential H is

$$\sum_{q'=1}^{2s} C_{q',n} w_{p',q',n}[\lambda_{q',n} - j(\sigma_{i,n}\cdot\hat{z})] \times \exp\{[\lambda_{q',n} - j(\sigma_{i,n}\cdot\hat{z})]nd/N\} = \sum_{q'=1}^{2s} C_{q',n+1} w_{p',q',n+1}[\lambda_{q',n+1} - j(\sigma_{i,n+1}\cdot\hat{z})] \times \exp\{[\lambda_{q',n+1} - j(\sigma_{i,n+1}\cdot\hat{z})]nd/N\}. \tag{24}$$

For the boundary between the Nth slab grating and region 3 ($z = d$), the boundary condition for tangential E is

$$\sum_{q'=1}^{2s} C_{q',N} w_{p',q',N} \exp\{[\lambda_{q',N} - j(\sigma_{i,N}\cdot\hat{z})]d\} = T_i, \tag{25}$$

and the boundary condition for tangential H is

$$\sum_{q'=1}^{2s} C_{q',N} w_{p',q',N}[\lambda_{q',N} - j(\sigma_{i,N}\cdot\hat{z})] \times \exp\{[\lambda_{q',N} - j(\sigma_{i,N}\cdot\hat{z})]d\} = -j(\mathbf{k}_{3i}\cdot\hat{z})\, T_i. \tag{26}$$

Equations (21)–(26) represent a total of $2(N+1)s$ equations. There are s unknown values each of R_i and T_i and $2s$ unknown values of $C_{q',n}$ for each slab grating. Thus the total number of unknowns is $2(N+1)s$, which is the same as the number of boundary-condition equations. If s values of i are retained in the analysis, then the calculations will yield s transmitted wave amplitudes (T_i) and s reflected wave amplitudes (R_i).

An efficient procedure to solve this large system of equations is to use a technique like Gauss elimination[21] applied successively to each boundary starting at the $z = 0$ input surface. By using this technique $N+1$ times in sequence, the s values of R_i and the s values of T_i may be obtained in a single pass on the last step. As depicted in Fig. 3, the boundary-condition equations are written as a matrix equation. The matrix is $2(N+1)s$ by $2(N+1)s$ and consists of the coefficients of $C_{q',n}$ (for $q' = 1$ to $2s$ and $n = 1$ to N), R_i (s values), and T_i (s values). For each slab grating, the boundary-condition equations for its two boundaries produces a $4s$ by $2s$ submatrix. Starting with the first ($n = 1$) slab grating (represented by upper-left-hand submatrix), a technique like Gauss elimination is applied to make all the elements of the lower half of the submatrix equal to zero. This reduces the system to $2Ns$ equations. Repeating this procedure on the next ($n = 2$) $4s$ by $2s$ submatrix reduces the system to $2(N-1)s$ equations. This process is continued until, after N steps, only $2s$ equations in the diffracted amplitudes R_i and T_i remain. These are then solved for R_i and T_i. At each step in this sequential process, a new set of coefficients of R_i is produced, as shown by the dashed box in Fig. 3. After N steps these coefficients have moved to the bottom of the matrix, and the final set of $2s$ equations in R_i and T_i are formed. This sequential procedure greatly reduces the storage and computational requirements for this type of problem. At each step, only a small $4s$ by $2s$ matrix is being treated, as opposed to the entire $2(N+1)s$ by $2(N+1)s$ matrix, where N might typically be 50–100.

When the amplitudes R_i and T_i are known, then the diffraction efficiencies (ratio of diffracted intensity to input intensity) may be directly determined. The diffraction efficiencies in regions 1 and 3 are

$$\mathrm{DE}_{1i} = \mathrm{Re}[(\mathbf{k}_{1i}\cdot\hat{z})/(\mathbf{k}_{10}\cdot\hat{z})]R_i R_i^* \tag{27}$$

and

$$\mathrm{DE}_{3i} = \mathrm{Re}[(\mathbf{k}_{3i}\cdot\hat{z})/(\mathbf{k}_{10}\cdot\hat{z})]T_i T_i^*. \tag{28}$$

The real part of the ratio of the propagation constants occurs when the time-average power-flow density is obtained by taking the real part of the complex Poynting vector. For

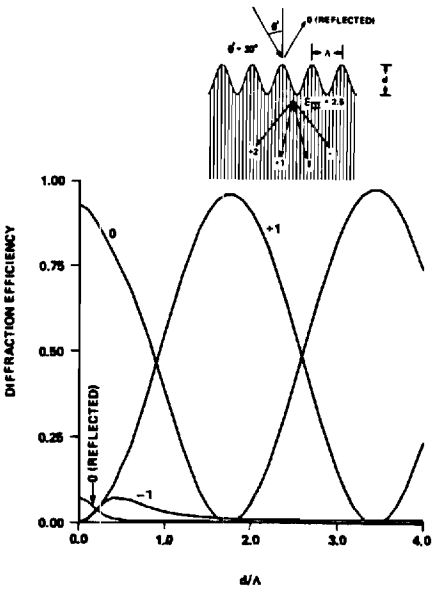


Fig. 4. Diffraction efficiency as a function of groove depth for a lossless sinusoidal surface-relief grating. Incidence is at first Bragg angle ($m = 1$), $\lambda = \Lambda$, $\epsilon_1 = 1.00$, and $\epsilon_{III} = 2.50$.

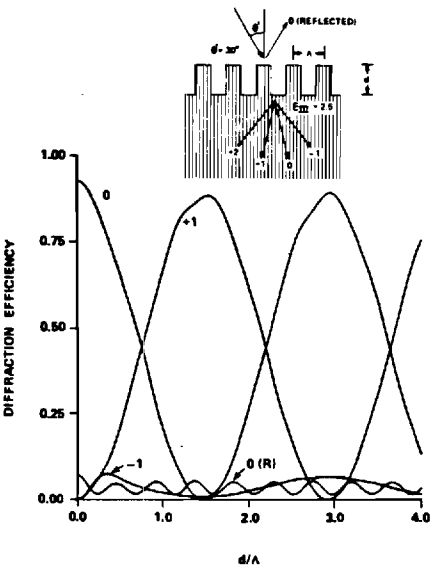


Fig. 5. Diffraction efficiency as a function of groove depth for a lossless square-wave surface-relief grating. Incidence is at first Bragg angle ($m = 1$), $\lambda = \Lambda$, $\epsilon_I = 1.00$, and $\epsilon_{III} = 2.50$.

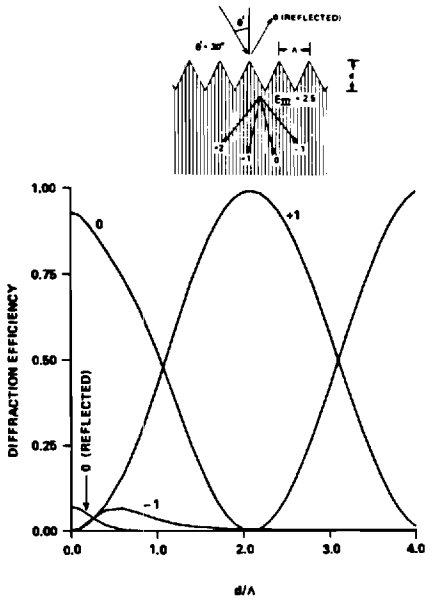


Fig. 6. Diffraction efficiency as a function of groove depth for a lossless triangular grating. Incidence is at first Bragg angle ($m = 1$), $\lambda = \Lambda$, $\epsilon_I = 1.00$, and $\epsilon_{III} = 2.50$.

lossless gratings the input power is conserved, and thus the sum of all the efficiencies for the propagating waves is unity. That is,

$$\sum_i (DE_{1i} + DE_{3i}) = 1. \tag{29}$$

This relationship may be used to verify the convergence of the numerical calculations.

4. RESULTS AND DISCUSSION

By using the method of solution described in Section 3, it is possible to calculate the diffraction efficiencies of dielectric surface-relief gratings to an arbitrary level of accuracy. The analysis contains no restrictions with respect to grating profile, groove depth, angle of incidence, or wavelength. Example results are presented in this section for sinusoidal (Fig. 4), square-wave (Fig. 5), triangular (Fig. 6), and sawtooth (Figs. 7 and 8) gratings. In each case the input wave is a plane wave with polarization perpendicular to the plane of incidence (H mode) and incident at the first Bragg angle ($m = 1$). For the cases presented, $\lambda = \Lambda$, and so the first Bragg angle is $\theta' = 30°$. Unlike with previous methods, large groove depths do not cause numerical instabilities or inaccuracies. Results for groove depths as deep as four grating periods ($d/\Lambda = 4$) are presented in Figs. 4–8. Previously published results have had maximum groove depths of $d/\Lambda = 0.60$ (Ref. 5), 1.0 (Ref. 14), and 2.4.[16] The input region has a relative permittivity of ϵ_I

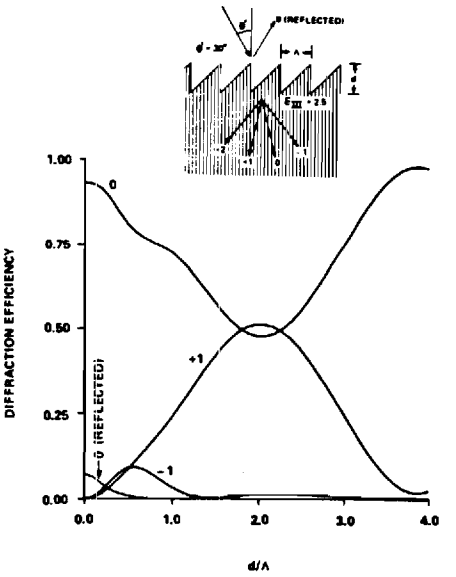


Fig. 7. Diffraction efficiency as a function of groove depth for a lossless sawtooth grating. Incidence is at first Bragg angle ($m = 1$) with the inclined surfaces of the sawtooth facing the incident beam. Also $\lambda = \Lambda$, $\epsilon_I = 1.00$, and $\epsilon_{III} = 2.50$.

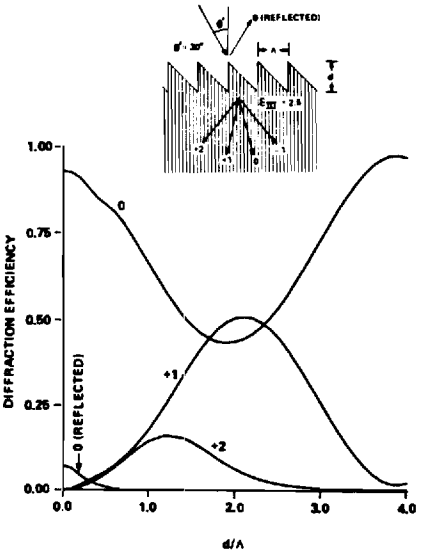


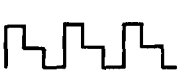
Fig. 8. Diffraction efficiency as a function of groove depth for a lossless sawtooth grating. Incidence is at first Bragg angle ($m = 1$) with the inclined surfaces of the sawtooth facing away from the incident beam. Also $\lambda = \Lambda$, $\epsilon_I = 1.00$, and $\epsilon_{III} = 2.50$.

= 1.00 (air), and the substrate a relative permittivity of ϵ_{III} = 2.50 (refractive index of 1.58).

With the exception of the zero-order wave, the reflected waves have diffraction efficiencies of less than 1%. The zero-order (specularly reflected) wave generally was found to decrease in intensity from the Fresnel reflection value with increasing groove depth. In the square-wave grating case, however, the zero-order reflected intensity is seen to oscillate with groove depth.

For the combination of relative permittivities used (ϵ_I = 1.00 and ϵ_{III} = 2.50), the transmitted first-order (i = +1) diffraction efficiency was found to be maximum for a groove depth in the range from 1.5 to 2 grating periods. For the sinusoidal, square-wave, and triangular grating profiles, the first-order diffraction efficiency was found to reach nearly 100% for the properly chosen groove depth. For the sawtooth gratings, however, the maximum first-order diffraction efficiency was only about 50%. A result that was consistently found in this work is the following: Dielectric gratings with profiles that are expressable as even functions are capable of high (greater than 85%) diffraction efficiency. Thus, if the grating profile is symmetric so that an x = 0 origin can be chosen so that $F(x)$ is an even function, then large first-order diffraction efficiencies are possible. The diffraction efficiencies calculated for any case are, of course, independent of the choice of the x = 0 origin. But if the grating has even symmetry, large diffraction efficiencies are possible. For grating profiles that lack even symmetry, such as the sawtooth profile, the maximum diffraction efficiencies are correspondingly less. Maximum diffraction efficiency results for several profiles are given in Table 1. Notice that the stairstep grating profile of even symmetry has a maximum first-order diffraction efficiency of 89.4%. However, the same stairstep grating with steps shifted, producing a profile of only odd symmetry, has a maximum first-order diffraction efficiency of only 67.7 or 71.8% (depending on the orientation of the Bragg angle). The lack of even symmetry obviously produces dephasing of the fundamental first-order diffracted beam and causes the diffraction efficiency not to reach a large value. Because the lack of even symmetry is less severe in the stairstep grating than in the sawtooth grating, the maximum efficiency for the stairstep is correspondingly larger. Further, it is noted from Table 1 that the choice of orientation of an asymmetric grating with respect to the Bragg angle does not appreciably affect the maximum possible diffraction efficiency. There is a modest trend among the first four gratings in Table 1 (each expressable as an even function) that indicates that higher maximum diffraction efficiencies are obtainable with sharper, more-pointed grating profiles. Thus, in the sequence of square-wave, stairstep, sinusoid, and triangular grating, the maximum diffraction efficiency steadily increases (from 88.5 to 99.0%). Again, however, the dominant factor controlling the maximum diffraction efficiency is the degree of even symmetry in the grating profile.

Table 1. Maximum Transmitted First-Order (i = +1) Diffraction Efficiencies for Various Grating Profiles[a]

	$(DE_{3,1})_{max}$	$\frac{d}{\Lambda}$
	99.0	2.10
	95.9	1.75
	89.4	1.60
	88.5	1.55
	71.8	1.62
	67.7	1.67
	51.0	2.10
	50.6	2.10

[a] Incidence is at first Bragg angle, θ' = 30°, ϵ_I = 1.00, and ϵ_{III} = 2.50. Diffraction efficiencies correspond to the value at the first peak of the $DE_{3,1}$ versus d/Λ curve.

The method of analysis presented in this paper is rigorous, and any arbitrary level of accuracy is attainable if needed. The method can be applied to any grating profile. Further, it is straightforward to add substrate and cover layers using this method of analysis if they are present.

5. NUMERICAL ACCURACY

In this section the accuracy, the convergence, and the stability of the numerical technique are discussed. Comparisons with previously published results from other techniques are also presented.

For all the numerical results presented, both the number of waves (space harmonics) included and the number of layers were increased until the changes in the calculated diffraction efficiencies were less than 10^{-8}. Typically 50–100 layers and 8–12 waves (values of i) were used. The algorithm described was *always* stable, and it converged in *every* case considered. The conservation of power was also always accurately achieved. In all our results the error in the conservation of power [Eq. (29)] was of the order of 10^{-12}.

Although the resulting system of simultaneous linear equations is $2(N+1)s$ by $2(N+1)s$, the technique described in this paper handles this large system by reducing it to the problem of converting a $4s$ by $2s$ system into upper-matrix form N times. This reduces the data-storage requirement by a factor of $(N+1)^2/3$ and reduces the execution time by a factor of N^2, where N is the number of layers. Without such savings, it would be extremely difficult to solve this large system of equations on even the largest of computers. By using the present method, a typical execution time for an 8-wave, 50-layer configuration is about 6 sec for each diffraction-efficiency point on a CDC 730/760 computer.

In the present work the rigorous coupled-wave approach is used to find the fields in each of the layers of the grating. Another possible technique is that of Lewis and Hessel,[18] which analyzes a periodic array of dielectric slabs. The coupled-wave approach is as accurate and as rigorous as the array-of-slabs technique. Both require finding the eigenvalues of a truncated characteristic or coupled-wave equation matrix. However, the coupled-wave approach is more suitable for computer implementation.

To show the accuracy and the effectiveness of the present

Table 2. Comparison of Diffraction Efficiencies Calculated by van den Berg[a] and Chang *et al.*[b] with Values Calculated in the Present Work[c]

Diffraction Efficiency	d/Λ	van den Berg	Chang *et al.*	Present Work
$DE_{1,0}$	0.005	0.203990	0.2037150	0.2037170
$DE_{1,1}$	0.005	0.000027	0.0000265	0.0000262
$DE_{3,-1}$	0.005	0.000074	0.0000817	0.0000820
$DE_{3,0}$	0.005	0.795207	0.7961250	0.7961205
$DE_{3,1}$	0.005	0.000056	0.0000550	0.0000543
$DE_{3,2}$	0.005	0.000000	1.3×10^{-9}	1.3×10^{-9}
$\sum$DE-1	0.005	-1.1×10^{-4}	3.6×10^{-6}	3.3×10^{-12}
$DE_{1,0}$	0.5	0.02418	0.0349913	0.0346051
$DE_{1,1}$	0.5	0.06367	0.0537149	0.0537065
$DE_{3,-1}$	0.5	0.11222	0.1693248	0.1690036
$DE_{3,0}$	0.5	0.39340	0.3492882	0.3484211
$DE_{3,1}$	0.5	0.34700	0.3307839	0.3323272
$DE_{3,2}$	0.5	0.05563	0.0618147	0.0619366
$\sum$DE-1	0.5	-3.9×10^{-3}	8.2×10^{-5}	1.7×10^{-12}

[a] Ref. 13.
[b] Ref. 16.
[c] The sinusoidal surface-relief grating is characterized by $\epsilon_I = 1.0$ and $\epsilon_{III} = 4.0$. The incident plane wave is characterized by $\lambda = \Lambda$ and $\theta' = 45°$.

Table 3. Comparison of Diffraction Efficiencies Calculated by Tremain and Mei[a] and Zaki[b] with Values Calculated in the Present Work[c]

Diffraction Efficiency	λ/Λ	Tremain and Mei	Zaki	Present Work
$DE_{3,0}$	0.750	0.202	0.16	0.15377
$DE_{3,1}$	0.750	0.547	0.50	0.57732
$DE_{3,2}$	0.750	0.241	0.29	0.20013
$\sum$DE-1	0.750	7×10^{-3}	-4×10^{-2}	1.35×10^{-12}
$DE_{3,0}$	0.941	0.301	0.30	0.30290
$DE_{3,1}$	0.941	0.647	0.62	0.63689
$DE_{3,2}$	0.941	0.050	0.03	0.05058
$\sum$DE-1	0.941	8×10^{-3}	-4×10^{-2}	1.34×10^{-12}
$DE_{3,0}$	1.030	0.339	0.33	0.34202
$DE_{3,1}$	1.030	0.643	0.555	0.62361
$DE_{3,2}$	1.030	0.023	0.01	0.02398
$\sum$DE-1	1.030	1×10^{-2}	9×10^{-2}	8.6×10^{-13}
$DE_{3,0}$	1.177	0.396	0.39	0.39857
$DE_{3,1}$	1.177	0.598	0.58	0.59051
$DE_{3,2}$	1.177	0.000	0.00	0.00000
$\sum$DE-1	1.177	6×10^{-3}	2×10^{-2}	1.36×10^{-12}
$DE_{3,0}$	1.471	0.493	0.49	0.49355
$DE_{3,1}$	1.471	0.497	0.47	0.49134
$DE_{3,2}$	1.471	0.000	0.00	0.00000
$\sum$DE-1	1.471	5×10^{-3}	2×10^{-2}	1.03×10^{-12}

[a] Ref. 14.
[b] Ref. 12.
[c] The sinusoidal surface-relief grating is characterized by $\epsilon_I = 1.0$, $\epsilon_{III} = 2.3104$, and $d/\Lambda = 1.1765$. The plane wave is incident at $\theta' = 36°$. The sum of the diffraction efficiencies includes other propagating waves even though these are not listed separately in this table.

technique, results obtained in the present work were compared with those results that are available in the literature. Our results show excellent agreement with the graphical results of Tremain and Mei[14] (their Figs. 8–10). Of course, this agreement is limited by the accuracy of the graphs. In Table 2, results calculated by the present method are compared with numerical results from Chang *et al.*[16] (their Table I). As can be seen, our results agree extremely well with theirs. However, in the present work, the deviation from the conservation of power ($\sim 10^{-12}$) is much smaller than that of either van den Berg[13] ($\sim 10^{-4}$) or Chang *et al.*[16] ($\sim 10^{-6}$). In Table 3 results obtained by the present rigorous coupled-wave approach are compared with those presented by Tremain and Mei[14] (their Table II). It is shown that the deviations from the conservation of power in the present results ($\sim 10^{-12}$) are much smaller than those of Zaki[12] ($\sim 10^{-2}$) or of Tremain and Mei[14] ($\sim 10^{-3}$). The small size of the residual error in the conservation of power in the present work provides additional strong evidence that the diffraction-efficiency results are similarly accurate. Thus it may be concluded that the rigorous coupled-wave method of analysis of surface-relief gratings is accurate, stable, reliable, and efficient.

ACKNOWLEDGMENT

This research was sponsored in part by the National Science Foundation and the Joint Services Electronics Program.

REFERENCES

1. R. Petit, ed., *Electromagnetic Theory of Gratings* (Springer-Verlag, Berlin, 1980).
2. M. Neviere, R. Petit, and M. Cadilhac, "About the theory of optical grating coupler-waveguide systems," Opt. Commun. **8**, 113–117 (1973).
3. S. T. Peng, H. L. Bertoni, and T. Tamir, "Analysis of periodic thin-film structures with rectangular profiles," Opt. Commun. **10**, 91–94 (1974).
4. S. T. Peng and T. Tamir, "Directional blazing of waves guided by asymmetrical dielectric gratings," Opt. Commun. **11**, 405–409 (1974).
5. S. T. Peng, T. Tamir, and H. L. Bertoni, "Theory of periodic dielectric waveguides," IEEE Trans. Microwave Theory Tech. **MTT-23**, 123–133 (1975).
6. D. Marcuse, "Exact theory of TE-wave scattering from blazed dielectric gratings," Bell Syst. Tech. J. **55**, 1295–1317 (1976).
7. K. Knop, "Rigorous diffraction theory for transmission phase gratings with deep rectangular grooves," J. Opt. Soc. Am. **68**, 1206–1210 (1978).
8. K. C. Chang and T. Tamir, "Simplified approach to surface-wave scattering by blazed dielectric gratings," Appl. Opt. **19**, 282–288 (1980).
9. W. Streifer, D. R. Scifres, and R. D. Burnham, "Analysis of grating-coupled radiation in GaAs:GaAlAs lasers and waveguides," IEEE J. Quantum Electron. **QE-12**, 422–428 (1976).
10. T. Tamir and S. T. Peng, "Analysis and design of grating couplers," Appl. Phys. **14**, 235–254 (1977).
11. A. R. Neureuther and K. Zaki, "Numerical methods for the analysis of scattering from nonplanar and periodic structures," Alta Freq. **38**, 282–285 (1969).
12. K. Zaki, "Numerical methods for the analysis of scattering from nonplanar periodic structures," Ph.D. thesis (University of California, Berkeley, Calif., 1969) (unpublished).
13. P. M. van den Berg, "Rigorous diffraction theory of optical reflection and transmission gratings," Ph.D. thesis (Delft University of Technology, Delft, The Netherlands, 1971), rep. no. 1971-16 (unpublished).
14. D. E. Tremain and K. K. Mei, "Application of the unimoment method to scattering from periodic dielectric structures," J. Opt. Soc. Am. **68**, 775–783 (1978).
15. K. C. Chang, "Surface-wave scattering by dielectric gratings with arbitrary profiles," Ph.D. thesis (Polytechnic Institute of New York, New York, 1979) (unpublished).
16. K. C. Chang, V. Shah, and T. Tamir, "Scattering and guiding of waves by dielectric gratings with arbitrary profiles," J. Opt. Soc. Am. **70**, 804–813 (1980).
17. M. G. Moharam and T. K. Gaylord, "Rigorous coupled-wave analysis of planar grating diffraction," J. Opt. Soc. Am. **71**, 811–818 (1981).
18. L. R. Lewis and A. Hessel, "Propagation characteristics of periodic arrays of dielectric slabs," IEEE Trans. Microwave Theory Tech. **MTT-19**, 276–286 (1971).
19. E.g., C. L. Liu and J. W. S. Liu, *Linear Systems Analysis* (McGraw-Hill, New York, 1975).
20. E.g., program EIGRF from the International Mathematics and Statistics Library (IMSL), Houston, Texas.
21. E.g., B. Carnahan, H. A. Luther, and J. O. Wilkes, *Applied Numerical Methods* (Wiley, New York, 1969).

Chain-matrix analysis of arbitrary-thickness dielectric reflection gratings

M. G. Moharam and T. K. Gaylord

School of Electrical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

Received August 12, 1981

A simple but rigorous chain-matrix analysis is used to determine the plane-wave diffraction efficiencies and angular selectivities of planar pure-reflection gratings (zero-slant angle) of arbitrary thickness, arbitrary-starting and arbitrary-ending conditions, and arbitrary incidence angle. The results from rigorous coupled-wave theory in the angular limit as the fringes become parallel to the surfaces are shown to approach these results in a particular average sense.

INTRODUCTION

Pure slab-reflection gratings have fringes parallel to the surfaces of the medium. It is due to their finite thicknesses that these structures are not strictly periodic. There may be a large number of grating periods present, but modal[1-8] and coupled-wave[9-15] grating theories based on the Floquet theorem require a truly periodic grating (an infinite number of periods). These theories may be applied in the angular limit as the slanted fringes of a general grating approach being parallel to the surface. However, for exact parallelism with the surface, a continuum of solutions is possible, depending on the number of periods, the starting conditions, and the ending conditions of the grating.

The chain-matrix method of solution[16,17] is often applied to the analysis of single-layer and multilayer thin-film coatings. However, a reflection grating with a continuous variation in permittivity may be represented as a large stack of thin homogeneous slabs.[18] If the individual planar slabs are allowed to be sufficiently thin, the resulting model can be made arbitrarily accurate. The chain matrix relates quantities at the input to the corresponding quantities at the output. In various forms, the chain matrix is also referred to as the characteristic matrix, the *ABCD* matrix, and the transmission matrix. This method of analysis is used here. It is rigorous for the case of pure-reflection gratings.

The purpose of this paper is (1) to present rigorously calculated plane-wave diffraction efficiency and angular selectivity results for dielectric reflection gratings showing the effect of grating thickness, the effect of fractional periods, and the effect of starting and ending conditions, and (2) to show that results from rigorous periodic-based theories approach, in the reflection grating limit, the average of the continuum of results from rigorous reflection grating theory.

THEORY

An obliquely incident plane wave illuminates a lossless dielectric reflection grating. The modulated region (region 2) extends from $z = 0$ to $z = d$. The regions outside the grating (regions 1 and 3) are homogeneous. The relative permittivity (dielectric constant) in the grating is of a general sinusoidal form given by

$$\epsilon(z) = \epsilon_2 + \Delta\epsilon \cos(Kz - \Phi), \qquad 0 \le z \le d, \tag{1}$$

where ϵ_2 is the average relative permittivity of region 2, $\Delta\epsilon$ is the amplitude of the grating, $K = 2\pi/\Lambda$, Λ is the grating period, and Φ is a phase constant that determines the starting condition for the grating ($\Phi = 0, \pi/2, \pi, 3\pi/2$ correspond, respectively, to cosine, sine, -cosine, -sine gratings). For this case of a planar-reflection grating with fringes exactly parallel to the surfaces, there are no higher-order waves: that is, the incident wave (region 1) is divided into a transmitted wave (region 3) and a diffracted (reflected) wave (region 1). There are no other waves for this grating configuration.

The planar-reflection grating given by Eq. (1) may be represented by a large number of sufficiently thin homogeneous slabs. The incident and reflected electric fields at the input of the ith slab may be related to those at the input of the $(i+1)$th slab by a chain matrix

$$\begin{pmatrix} E_{\text{inc},i} \\ E_{\text{refl},i} \end{pmatrix} = \begin{pmatrix} A_i & B_i \\ B_i^* & A_i^* \end{pmatrix} \begin{pmatrix} E_{\text{inc},i+1} \\ E_{\text{refl},i+1} \end{pmatrix}. \tag{2}$$

In this type of formulation, the effect of having multiple slab layers together is represented simply by multiplying their respective chain matrices together and then multiplying that product by the chain matrix of region 3 ($z > d$). Thus, for a reflection grating that is divided into n thin slabs, the governing equation is

$$\begin{pmatrix} 1 \\ E_r \end{pmatrix} = \prod_{i=1}^{n} \begin{pmatrix} A_i & B_i \\ B_i^* & A_i^* \end{pmatrix} \times \frac{1}{\tau_{n+1}} \begin{pmatrix} 1 & r_{n+1} \\ r_{n+1} & 1 \end{pmatrix} \begin{pmatrix} E_t \\ 0 \end{pmatrix}, \tag{3}$$

where E_r and E_t are the reflected and transmitted normalized electric-field amplitudes, respectively. For a lossless dielectric layer, the matrix elements are

$$A_i = (1/\tau_i)\exp[j2\pi(\epsilon_{2,i})^{1/2} d_i \cos\theta_{2,i}/\lambda], \tag{4}$$

$$B_i = r_i A_i^*, \tag{5}$$

where λ is the free-space wavelength and $\epsilon_{2,i}$, $\theta_{2,i}$, and d_i are the relative permittivity, the angle of refraction, and the thickness of the ith layer. The amplitude reflection and

transmission coefficients are

$$r_i = (Z_i - Z_{i-1})/(Z_i + Z_{i-1}), \qquad \tau_i = 1 + r_i, \tag{6}$$

where Z_i is the normalized characteristic impedance of the ith layer. For H-mode polarization (electric field perpendicular to plane of incidence),

$$\begin{aligned} Z_0 &= 1, \quad \text{region 1}, \\ Z_i &= \cos\theta/[(\epsilon_{2,i}/\epsilon_1) - \sin^2\theta]^{1/2}, \quad 1 \le i \le n, \\ Z_{n+1} &= \cos\theta/[(\epsilon_3/\epsilon_1) - \sin^2\theta]^{1/2}, \quad \text{region 3}, \end{aligned} \tag{7}$$

and for E-mode polarization (electric field parallel to plane of incidence),

$$\begin{aligned} Z_0 &= 1, \quad \text{region 1}, \\ Z_i &= (\epsilon_1/\epsilon_{2,i})[(\epsilon_{2,i}/\epsilon_1) - \sin^2\theta]^{1/2}/\cos\theta, \quad 1 \le i \le n, \\ Z_{n+1} &= (\epsilon_1/\epsilon_3)[(\epsilon_3/\epsilon_1) - \sin^2\theta]^{1/2}/\cos\theta, \quad \text{region 3}, \end{aligned} \tag{8}$$

where θ is the angle of incidence in region 1. The diffraction efficiency (fraction of power reflected) is simply $DE = E_r E_r^*$.

RESULTS AND DISCUSSION

Since much of the literature describes the simple case of the same average permittivity inside and outside the grating ($\epsilon_1 = \epsilon_2 = \epsilon_3$), the situation is also used here. This case applies directly, for example, in integrated optics. However, this is not an essential assumption, and the theory applies in general.

At the input surface ($z = 0$), the grating may start with any value of relative permittivity. The permittivity may be continuous at $z = 0$, as in the cases of the +sine and −sine gratings, or may be discontinuous by a maximum amount, as in the +cosine and −cosine gratings. The starting condition for the grating is determined by the value of Φ in Eq. (1). There is a continuum of starting conditions given by the range $0 \le \Phi \le 2\pi$.

Typical reflected intensities are shown in Fig. 1 for H-mode polarization for a thickness of 3.125 grating periods and $\rho = 20$, where ρ, the Bragg-regime parameter, is given by $\rho = 2\lambda^2/\Lambda^2\Delta\epsilon$. The input wave is incident at the Bragg angle. That is, the Bragg condition $\lambda/(\epsilon_1)^{1/2} = 2\Lambda\cos\theta$ is satisfied. Thus, for the present case, $\rho = 8\cos^2\theta(\epsilon_1/\Delta\epsilon)$. For the particular grating thickness and modulation of Fig. 1, the $\Phi = 9\pi/8$ grating (between a −cosine grating and a −sine grating) has the largest diffraction efficiency (58.8%). For the same thickness and modulation the $\Phi = 5\pi/8$ grating (between a +sine grating and a −cosine grating) and the $\Phi = 13\pi/8$ grating (between a −sine grating and a +cosine grating) both have the lowest diffraction efficiency (54.3%). Gratings with all other starting conditions (values of Φ) have efficiencies between these two values. This continuum of solutions is represented by the vertical line in Fig. 1. For any thickness and modulation, the maximum diffraction efficiency is found to occur for the grating with

$$\Phi = [1 + (d/\Lambda)\text{modulo } 1]\pi \tag{9}$$

for $0 < (d/\Lambda)\text{modulo } 1 \le 0.5$ and

$$\Phi = [\pm(1/2) + (d/\Lambda)\text{modulo } 1]\pi \tag{10}$$

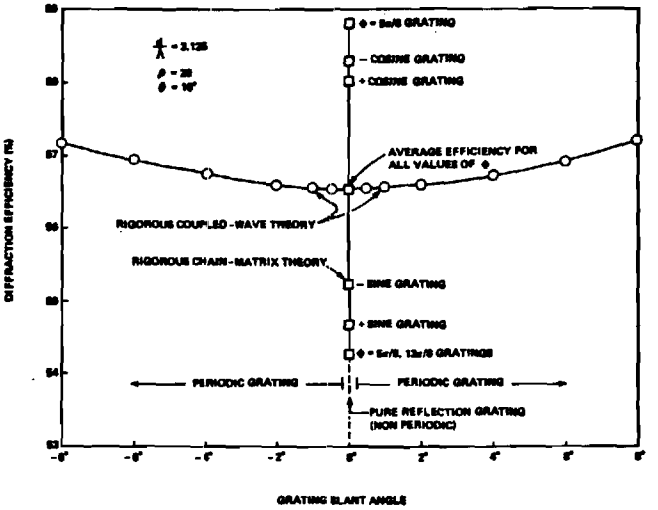


Fig. 1. Rigorously calculated total on-Bragg-angle diffraction efficiencies for H-mode polarization for dielectric reflection gratings with $\epsilon_1 = \epsilon_2 = \epsilon_3$. The results from rigorous coupled-wave theory for slanted reflection gratings are shown to approach the average of the continuum of solutions for the unslanted pure-reflection gratings.

for $0.5 < (d/\Lambda)\text{modulo } 1 \le 1.0$. For example, if $d/\Lambda = 5.5$, the maximum efficiency occurs for $\Phi = 3\pi/2$ (a −cosine grating). If $d/\Lambda = 7.75$, the maximum efficiency occurs for gratings with $\Phi = \pi/4$ and $\Phi = 5\pi/4$. The ending condition for a grating is determined by its thickness. By the reciprocity principle the +sine and −sine gratings produce the same efficiency for thicknesses that are an integer number of periods. Likewise, the +cosine and −cosine gratings have the same efficiency for thicknesses that are an odd-integer multiple of a half-period.

Also plotted in Fig. 1 is the total diffracted intensity as calculated by rigorous coupled-wave theory. This theory applies to periodic gratings with slanted fringes. The grating slant angle may approach (but not equal) zero in this theory. A grating slant angle of zero corresponds to a nonperiodic pure-reflection grating. The transition of the grating slant angle from near zero to exactly zero involves a singularity. The total diffraction efficiency changes from a single-valued function to a multiple-valued function (a continuum of solutions), as shown in Fig. 1. An important result of this work is that the rigorous coupled-wave theory in the limit of vanishingly small slant angle predicts a total diffraction efficiency equal to the average of the pure-reflection grating efficiencies with respect to the entire range of starting conditions ($0 \le \Phi \le 2\pi$). That is, for a given thickness and given modulation pure-reflection grating, the average diffraction efficiency for all possible starting conditions is equal to the diffraction efficiency for a slightly slanted reflection grating. In the case of a slanted-fringe reflection grating, the permittivity is periodic along the grating surfaces ($z = 0, d$) even if the slant is small. Only for the exactly unslanted case is the permittivity equal to a constant over all the grating surfaces. It seems intuitively correct that a slightly slanted reflection grating behaves like the average of all possible reflection gratings of the same thickness and modulation. The slanted grating is periodic along the $z = 0$ surface, at some points starting at a peak (like a cosine grating), at some points starting at a minimum (like a −cosine grating), etc.

The diffraction efficiencies of the +cosine and +sine gratings as calculated by the rigorous chain-matrix theory for H-mode polarization for the cases of $\epsilon_1 = \epsilon_2 = \epsilon_3$ are shown in Fig. 2 as a deviation from the average value for all possible starting conditions. This average value is within $\pm 1/\rho^2$ (which in this case is ±0.25%) of the diffraction efficiency predicted by $DE = \tanh^2[\pi \Delta\epsilon d/2\lambda(\epsilon_1)^{1/2} \cos\theta]$ from Kogelnik.[9] As can be seen from this figure, a grating thickness that is a multiple of $\Lambda/4$ produces a small (or zero) deviation of the diffraction efficiency from the average. The largest deviations from the average occur for thicknesses that are odd multiples of $\Lambda/8$. A plot of the deviations for the −cosine and −sine gratings is similar to Fig. 2 and has all the basic features of Fig. 2. When the grating has a thickness that is an integer multiple of Λ, the diffraction efficiencies of the ±sine gratings are exactly the same. For this case, there are small deviations in efficiency between the +cosine grating, the −cosine grating, and the ±sine gratings. As the thickness of the reflection grating increases, the deviations from the average diffraction efficiency, as shown in Fig. 2, increase as d/Λ increases from zero to about 3, and then the deviation oscillations decrease in amplitude as d/Λ increases further. For this case of $\rho = 20$ and for a grating thickness greater than 10 periods, the presence of fractional periods produces essentially negligible variations from the case of integer period thickness. The diffraction efficiency deviation from the average value (or equivalently the deviation from the rigorous coupled-wave theory result) does not exceed approximately ±2.5% efficiency for any thickness. Thus the results from rigorous coupled-wave theory with the grating slant angle approaching zero are within ±2.5% of the rigorous pure-reflection grating results regardless of the thickness and the starting conditions. In addition, the average value of the diffraction efficiency for all starting conditions is within $\pm 1/\rho^2$ of the value predicted by Kogelnik's theory.[9] Therefore, in this case, the deviations (±2.5%) that are due to starting conditions are an order of magnitude greater than the error (±0.25%) that is due to the neglect of second derivatives of field amplitudes and neglect of boundary diffraction in the Kogelnik coupled-wave formulation.

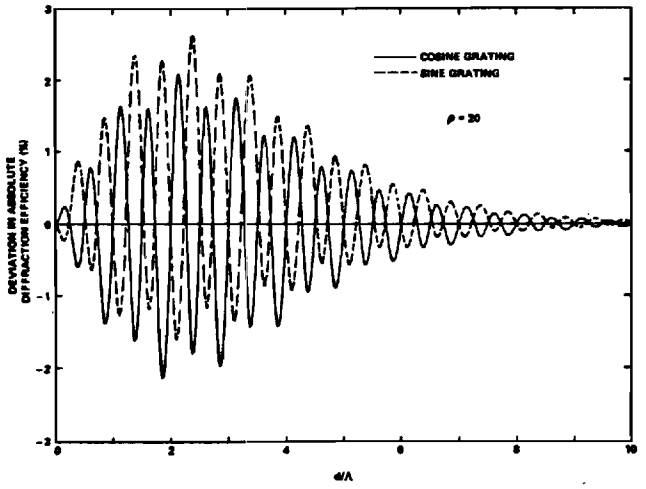


Fig. 2. Deviation of the rigorously calculated on-Bragg-angle diffraction efficiencies for H-mode polarization for +cosine and +sine gratings from the average of the continuum of diffraction efficiencies for pure-reflection gratings as a function of thickness.

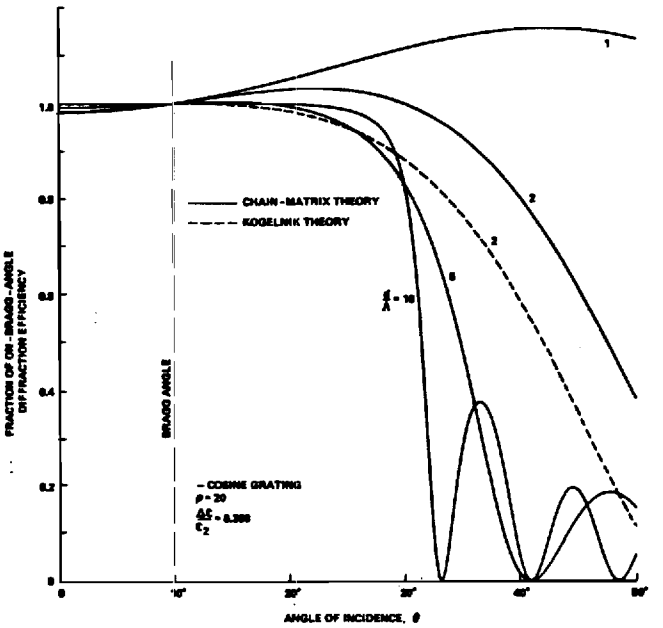


Fig. 3. Angular selectivity characteristics (fraction of on-Bragg-angle diffracted power that is reflected as a function of angle of incidence) for H-mode polarization for pure-reflection gratings of various thicknesses.

Typical plots of chain-matrix-calculated angular selectivity responses of pure-reflection gratings (−cosine gratings) are shown in Fig. 3. The diffraction efficiency is normalized to the value at the Bragg angle (10° in this example case). The angular selectivity is symmetric about normal incidence. The diffracted (reflected) power remains nearly constant from normal incidence to the Bragg angle. Beyond the Bragg angle the diffracted power decreases, thus producing the well-known angular selectivity feature of thick gratings.[9] For gratings of only two periods' thickness, the large Fresnel-reflection component may cause the total reflected power to increase initially for angles greater than the Bragg angle. As the thickness of the grating increases, the angular selectivity is enhanced. That is, to achieve a large fraction of reflected power, the input wave must be incident in an angular corridor of decreasing width as the grating thickness increases. For comparison, the corresponding angular selectivity from Kogelnik's approximate two-wave coupled-wave theory is also plotted for the case of a grating that is two periods thick. This approximate theory predicts a smaller angular selectivity width than that from rigorous theory. The approximate Kogelnik theory also does not predict the initial rise in reflected power beyond the Bragg angle that is due to the Fresnel component of the reflection. Angular selectivity plots for ±sine and +cosine gratings are similar in general form to the −cosine grating case shown in Fig. 3.

SUMMARY

A rigorous chain-matrix method of analysis has been used to calculate the diffraction characteristics of pure- (unslanted) reflection gratings. Since pure-reflection gratings diffract (reflect) only a single order, the diffraction efficiency of this order is a complete description of the diffraction process. The analysis has been applied to reflection gratings of arbitrary

thickness, arbitrary starting and ending conditions, and arbitrary incidence angle. For the same average permittivity inside and outside the grating, a variation in the starting conditions of the grating (e.g., ±cosine and ±sine grating) produces approximately ±2.5% maximum change in the diffraction efficiency from the average value for a grating with $\rho = 20$. The results from rigorous coupled-wave theory, in the angular limit of the fringes' becoming parallel to the surface, have been shown to approach the average value of the pure-reflection grating diffraction efficiencies for all possible starting conditions.

The rigorous chain-matrix method has been applied here to sinusoidal grating profiles with the same average permittivity inside and outside the grating ($\epsilon_1 = \epsilon_2 = \epsilon_3$). However, other grating profiles and other surrounding permittivities may be treated by using the analysis presented without modification.

REFERENCES

1. C. B. Burckhardt, "Diffraction of a plane wave at a sinusoidally stratified dielectric grating," J. Opt. Soc. Am. **56**, 1502–1509 (1966).
2. T. Tamir and H. C. Wang, "Scattering of electromagnetic waves by a sinusoidally stratified half-space: I. Formal solution and analytic approximations," Can. J. Phys. **44**, 2073–2094 (1966).
3. T. Tamir, "Scattering of electromagnetic waves by a sinusoidally stratified half-space: II. Diffraction aspects at the Rayleigh and Bragg wavelengths," Can. J. Phys. **44**, 2461–2494 (1966).
4. L. Bergstein and D. Kermisch, "Image storage and reconstruction in volume holography," Proc. Symp. Modern Opt. **17**, 655–680 (1967).
5. S. T. Peng and E. S. Cassedy, "Scattering of light waves at boundaries to parametrically modulated media," Proc. Symp. Modern Opt. **17**, 299–342 (1967).
6. F. G. Kaspar, "Diffraction by thick periodically stratified gratings with complex dielectric constant," J. Opt. Soc. Am. **63**, 37–45 (1973).
7. R. S. Chu and T. Tamir, "Wave propagation and dispersion in space-time periodic media," Proc. IEE **119**, 797–806 (1972).
8. R. S. Chu and J. A. Kong, "Modal theory of spatially periodic media," IEEE Trans. Microwave Theory Tech. **MTT-25**, 18–24 (1977).
9. H. Kogelnik, "Coupled wave theory for thick hologram gratings," Bell Syst. Tech. J. **48**, 2909–2947 (1969).
10. R. S. Chu and T. Tamir, "Guided-wave theory of light diffraction by acoustic microwaves," IEEE Trans. Microwave Theory Tech. **MTT-18**, 486–504 (1970).
11. D. L. Jaggard and C. Elachi, "Floquet and coupled-wave analysis of higher-order Bragg coupling in a periodic medium," J. Opt. Soc. Am. **66**, 674–682 (1976).
12. W. R. Klein and B. D. Cook, "Unified approach to ultrasonic light diffraction," IEEE Trans. Sonics Ultrason. **SU-14**, 123–134 (1967).
13. R. Magnusson and T. K. Gaylord, "Analysis of multiwave diffraction of thick gratings," J. Opt. Soc. Am. **67**, 1165–1170 (1977).
14. M. G. Moharam and T. K. Gaylord, "Coupled-wave analysis of reflection gratings," Appl. Opt. **20**, 240–244 (1981).
15. M. G. Moharam and T. K. Gaylord, "Rigorous coupled-wave analysis of planar grating diffraction," J. Opt. Soc. Am. **71**, 811–818 (1981).
16. M. Born and E. Wolf, *Principles of Optics*, 6th ed. (Pergamon, Oxford, 1980), pp. 55–70.
17. R. E. Collin, *Field Theory of Guided Waves* (McGraw-Hill, New York, 1960), pp. 79–94.
18. W. W. Rigrod, "Diffraction efficiency of nonsinusoidal Bragg reflection gratings," J. Opt. Soc. Am. **64**, 97–99 (1974); Erratum, J. Opt. Soc. Am. **64**, 895 (1974).

Comments on analyses of reflection gratings

M. G. Moharam and T. K. Gaylord

School of Electrical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

Received August 30, 1982

The descriptions of diffraction by pure reflection gratings (fringes parallel to surfaces) and by reflection gratings with slanted fringes are clarified. The rigorous coupled-wave analysis, the chain-matrix analysis, and the new method presented by Zylberberg and Marom [J. Opt. Soc. Am. **73**, 392 (1983)] are compared and their regions of applicability determined. Zylberberg and Marom refer to their method as a coupled-wave analysis, a description that is inconsistent with previous use of this terminology.

INTRODUCTION

Because of its widespread importance, grating diffraction has been the subject of considerable investigation. Numerous methods of analysis have been applied to this problem. In 1956, Phariseau[1] introduced coupled-wave theory to analyze acoustic gratings. In 1967, Klein and Cook[2] used a multiwave first-order coupled-wave analysis to describe transmission grating diffraction. In 1969, Kogelnik[3] presented a general two-wave first-order coupled-wave analysis of both transmission and reflection gratings. This treatment included absorption and phase gratings, E and H polarizations, and slanted-fringe gratings. In 1977, Kong[4] presented a two-wave second-order coupled-wave analysis of transmission gratings. In 1981, a rigorous (without approximation) coupled-wave analysis of transmission and reflection gratings was introduced by Moharam and Gaylord.[5] The last-named method has generated considerable interest because it is an exact formulation, and calculations can be performed with any desired level of accuracy. However, in all coupled-wave analyses the field inside the grating is expanded in terms of space-harmonic (Floquet) components.

An example alternative method that was introduced by Marcuse[6] involves the expansion of the space-harmonic amplitudes in a Fourier series. Since the space-harmonic amplitudes are not periodic in the grating, the period used in the expansion must be chosen suitably larger than the grating thickness, and a relatively large number of terms must be used to obtain good convergence. Zylberberg and Marom, in Appendix A of Ref. 7, analyze the pure-reflection-grating case by correctly expanding the field that is present in a Fourier series in α, the phase-shift parameter that describes the starting condition of the grating. They refer to this method of analysis as rigorous coupled-wave theory, but this terminology is not consistent with the previous use of this expression.

REFLECTION-GRATING DIFFRACTION

Reflection gratings with slanted fringes ($\phi \neq 0$; see Fig. 1 of Ref. 7) were previously analyzed by using rigorous coupled-wave theory.[5,8] Pure reflection gratings with fringes parallel to the surface ($\phi \equiv 0$) have also been analyzed by the chain-matrix method,[9,10] in which the grating is represented as a large stack of thin homogeneous slabs, and by a new method presented by Zylberberg and Marom.[7] In previous rigorous coupled-wave analyses, Moharam and Gaylord[5,8] were careful to indicate that they were treating the slanted-fringe-grating case and not the pure-reflection-grating case (contrary to the interpretation of this work presented in Ref. 7). This is evident from the following quoted statements from the first paper on rigorous coupled-wave analysis (p. 241 of Ref. 8): "It is important to note that by taking ϕ approaching, but not equal to, zero, all of the individually calculated values of β_i converge toward the same value. That is, all of the reflected waves (R_i) and all of the transmitted waves (T_i) have virtually the same propagation constant but are still individually distinguishable. With an understanding of this limiting process, ϕ can be taken to be equal to zero in the calculations."

There are a number of important distinctions between the slanted- ($\phi \neq 0$) and unslanted- ($\phi \equiv 0$) fringe reflection-grating cases. First, when the grating slant angle is exactly zero, there is a continuum of solutions possible depending on the grating phase-shift parameter (α). For example, a sine grating $\epsilon(z) = \epsilon_0 + \epsilon_1 \sin Kz$ and a cosine grating $\epsilon(z) = \epsilon_0 + \epsilon_1 \cos Kz$, with all other parameters held constant, would typically have different diffraction efficiencies. Here $\epsilon(z)$ is the relative permittivity, ϵ_0 is its average value, ϵ_1 is the amplitude of its sinusoidal component, and K is the magnitude of the grating vector ($K = 2\pi/\Lambda$, where Λ is the grating period), and the grating extends from the plane $z = 0$ to the plane $z = d$. However, for $\phi \neq 0$ the continuum of solutions disappears and only a single solution is obtained. This is because the grating is now periodic along the boundary, and all possible starting conditions from $\epsilon_0 - \epsilon_1$ to $\epsilon_0 + \epsilon_1$ are presented to the incident plane wave. In this slanted case the value of α does not affect the diffraction-efficiency results in any way.

Second, in the limit of unslanted fringes ($\phi \equiv 0$), the field inside the grating is not periodic along the boundary. For any value of z inside the grating, the field is composed of two plane waves, one with a component in the positive z direction and one with a component in the negative z direction. This field is phase matched to only two waves outside the grating. These are the transmitted wave (T) and the reflected wave (R), as shown in Fig. 1. In this limit, the pure reflection grating may be analyzed in a straightforward manner without field expansion by using the chain-matrix method.[9,10] For

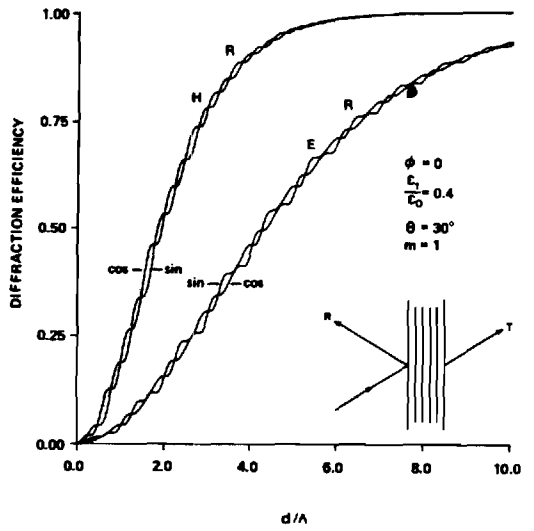


Fig. 1. Rigorously calculated first-Bragg-angle ($m = 1$) reflected diffraction efficiencies for an unslanted-fringe ($\phi \equiv 0$) pure reflection grating for H-mode polarization (electric field perpendicular to plane of incidence) and for E-mode polarization (electric field in plane of incidence). Results for a sinusoidal relative permittivity [$\epsilon(z) = \epsilon_0 + \epsilon_1 \sin Kz$] and a cosinusoidal relative permittivity [$\epsilon(z) = \epsilon_0 + \epsilon_1 \cos Kz$] are given. The average relative permittivity is the same inside and outside the grating.

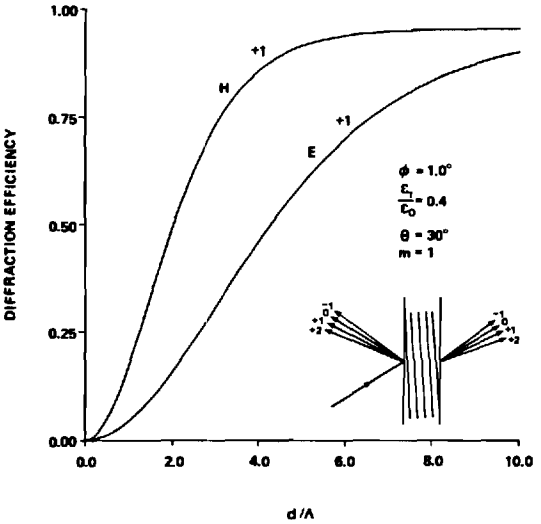


Fig. 2. Rigorously calculated first-Bragg-angle reflected diffraction efficiencies for a slanted-fringe reflection grating. There is a 1.0° angle between the grating fringes and the grating surfaces ($\phi = 1.0°$). The grating modulation level and the Bragg angle are the same as in Fig. 1. Notice the disappearance of the oscillatory behavior of the diffraction efficiency for fringes not parallel to the grating surfaces. The diffraction efficiencies of all the reflected orders (other than the +1 order) and all the transmitted orders (other than the zero order) are less than 2%.

the unslanted-fringe, pure-reflection-grating case, the diffraction efficiency tends to oscillate with increasing grating thickness. This is illustrated by the rigorous calculations in Fig. 1 for the cases of a sine grating [$\epsilon(z) = \epsilon_0 + \epsilon_1 \sin Kz$] and a cosine grating [$\epsilon(z) = \epsilon_0 + \epsilon_1 \cos Kz$]. The input boundary for the grating is at $z = 0$. This oscillatory behavior was shown in Ref. 10 and is shown again by Zylberberg and Marom.[7] However, for even slightly slanted fringes, the grating diffraction efficiency becomes a smooth, monotonic function of thickness. This is shown with rigorous calculations in Fig. 2 for a grating with exactly the same parameters as that in Fig. 1 except for the slightly slanted fringes ($\phi = 1.0°$). The monotonic diffraction-efficiency characteristic curve remains until the fringes become exactly parallel to the surface, and then oscillations are observed.

Third, for unslanted-fringe reflection gratings, the average diffraction efficiency over the entire range of starting conditions ($0 \le \alpha \le 2\pi$) is the same as that predicted by rigorous coupled-wave theory in the limit as ϕ approaches zero. This was first shown in Ref. 10, and now the details of the proof are provided by Zylberberg and Marom[7] in their Appendix B.

DISCUSSION

In the body of their accompanying paper, Zylberberg and Marom[7] analyze slanted-fringe reflection gratings using true rigorous coupled-wave theory. Since the phase-shift parameter α does not affect the diffraction-efficiency results, this formulation duplicates the previously published rigorous coupled-wave analysis.[5] For slanted-fringe gratings and plane-wave incidence, the value of α may be chosen arbitrarily, and choosing it is equivalent to selecting the origin for the x axis. No matter how α is selected, the resulting diffraction efficiencies are unchanged, as was pointed out previously in Ref. 10.

In Appendix A of their paper, Zylberberg and Marom present a new, interesting method for analyzing pure reflection gratings ($\phi \equiv 0$). In this case of zero slant, the coupled-wave expansion in terms of space harmonics is no longer an expansion in terms of space harmonics. In this case, all the space-harmonic components inside the grating collapse into one field (composed of two plane waves) that is phase matched to the single transmitted and to the single reflected fields. The expansion of the field (which is periodic along the boundary with the boundary periodicity of the grating) in terms of space-harmonic components for $\phi \neq 0$ is replaced by an expansion of this field in terms of the phase-shift parameter when $\phi \equiv 0$. The analysis is no longer a coupled-wave analysis. In all coupled-wave grating analyses, the field is periodic along the boundary; it is expanded in *space harmonic components* inside the grating, and these correspond to diffracted orders outside the grating. In the expansion used in Appendix A of Ref. 7 to perform all the calculations presented, the integer i is not over diffracted "orders," as shown in Figs. 1 and 4 of Ref. 7 (since there is only one diffracted "order" for $\phi \equiv 0$) but is over *terms* in the Fourier expansion of the electric field with a periodicity in the phase-shift parameter. The S_i quantities in Appendix A of Ref. 7 do not correspond to the S_i space-harmonic fields used in the body of that paper and in other coupled-wave analyses.[3,5,8,11] This represents an unfortunate mixing of notation. Note that, although Eqs. (12) and (A.12) for $\phi \equiv 0$ are mathematically identical, the quantities represented are not the same.

An important difference between all coupled-wave analyses (expansion in terms of space harmonics) and the expansion of Zylberberg and Marom (Fourier expansion in terms of phase-shift parameter) relates to conservation of energy. By the nature of the coupled-wave expansion, conservation of energy for a lossless grating is automatically incorporated in the analysis. This has been proved for the rigorous coupled-wave analysis,[5] multiwave first-order coupled-wave analysis,[11] two-wave second-order coupled-wave analysis,[4]

two-wave first-order coupled-wave analysis,[3] and Raman–Nath theory[12–14] (which is multiwave first-order coupled-wave analysis in which dephasing from the Bragg angle is ignored). In coupled-wave analyses, conservation of energy is obtained independently of the number of orders retained in the analysis. The accuracy of the diffraction efficiency of a particular order depends on the total number of orders retained in the analysis, but convergence is rapid with increasing number of orders retained.

By comparison, the Fourier expansion in terms of the phase-shift parameter used by Zylberberg and Marom requires a large number of terms (rather than "orders") to approach both energy-conservation and diffraction-efficiency convergence. This is shown in Figs. 4 and 5 of Ref. 7. In contrast, the convergence of the space-harmonic amplitudes in the coupled-wave analyses has consistently been found to be rapid, requiring relatively few orders to be retained.

For slanted-fringe ($\phi \neq 0$) reflection gratings, the number of space-harmonic components is infinite even though these decrease in amplitude rapidly with increasing order. By comparison, unslanted-fringe ($\phi \equiv 0$) reflection gratings have a field that is not periodic along the boundary and thus cannot be expanded in space-harmonic components. In the case of slanted-fringe reflection gratings, each space-harmonic component inside the grating contributes a separate distinct forward- and backward-diffracted order, as shown in Fig. 2. Each order has its own individual diffraction efficiency. As ϕ approaches zero, the forward-diffracted orders angularly collapse, approaching a single transmitted wave, and the backward-diffracted orders angularly collapse, approaching a single reflected wave, as shown in Fig. 2. In the practical case of finite beams, these diffracted beams may or may not be sufficiently separated to be measured individually. If they are separable, the individual diffraction efficiencies may be measured. If all the diffracted orders fall on the detector, then the sum of the individual *intensities* will be measured. This will apply when ϕ approaches infinitesimally close to zero but does not reach zero.

For unslanted-fringe reflection gratings, a field expansion is not necessary, as has been shown by the chain-matrix analysis method.[9,10] However, if the field is expanded in the pure-reflection-grating case (as has been done by Zylberberg and Marom[7]), it is correct to sum all the individual Fourier *amplitude components* (as opposed to intensities) to determine the diffraction efficiency. This has been correctly done by Zylberberg and Marom.[7] However, the summing of intensities of the diffracted orders for slanted-fringe reflection gratings to obtain the total diffraction efficiency, as done in Ref. 8, is correct for the case of ϕ infinitesimally close to zero and is not "erroneous," as stated in Ref. 7.

As has been discussed, pure reflection gratings may be analyzed by the rigorous chain-matrix analysis method[10] (not mentioned in Ref. 7). This method has the strong advantages that (1) an expansion of the field is not required, (2) no eigenvalue problem need be solved, (3) numerical implementation is simple, (4) numerical computation is very rapid (typically 25 times faster than a rigorous coupled-wave computation), (5) E-mode and H-mode polarization may be treated with equal ease, and (6) any grating profile (not just sinusoidal) may be analyzed. These advantages are not possessed by the pure-reflection-grating analysis method described in Ref. 7.

ACKNOWLEDGMENT

The authors wish to thank the editor of this journal for inviting this response.

REFERENCES

1. P. Phariseau, "On the diffraction of light by progressive supersonic waves," Proc. Ind. Acad. Sci. Sect. A **44**, 165–170 (1956).
2. W. R. Klein and B. D. Cook, "Unified approach to ultrasonic light diffraction," IEEE Trans. Sonics Ultrason. **SU-14**, 123–134 (1967).
3. H. Kogelnik, "Coupled-wave theory for thick hologram gratings," Bell Syst. Tech. J. **48**, 2909–2947 (1967).
4. J. A. Kong, "Second-order coupled-mode equations for spatially periodic media," J. Opt. Soc. Am. **67**, 825–829 (1977).
5. M. G. Moharam and T. K. Gaylord, "Rigorous coupled-wave analysis of planar-grating diffraction," J. Opt. Soc. Am. **71**, 811–818 (1981).
6. D. Marcuse, "Exact theory of TE-wave scattering from blazed dielectric gratings," Bell Syst. Tech. J. **55**, 1295–1317 (1976).
7. Z. Zylberberg and E. Marom, "Rigorous coupled-wave analysis of pure reflection gratings," J. Opt. Soc. Am. **73**, 392–398 (1983).
8. M. G. Moharam and T. K. Gaylord, "Rigorous coupled-wave analysis of reflection gratings," Appl. Opt. **20**, 240–244 (1981).
9. W. W. Rigrod, "Diffraction efficiency of nonsinusoidal Bragg reflection gratings," J. Opt. Soc. Am. **64**, 97–99 (1974); Erratum, J. Opt. Soc. Am. **64**, 895 (1974).
10. M. G. Moharam and T. K. Gaylord, "Chain-matrix analysis of arbitrary-thickness dielectric reflection gratings," J. Opt. Soc. Am. **72**, 187–190 (1982).
11. R. Magnusson and T. K. Gaylord, "Analysis of multiwave diffraction of thick gratings," J. Opt. Soc. Am. **67**, 1165–1170 (1977).
12. C. V. Raman and N. S. N. Nath, "The diffraction of light by high frequency sound waves: Part I," Proc. Ind. Acad. Sci. Sect. A **2**, 406–412 (1935).
13. C. V. Raman and N. S. N. Nath, "The diffraction of light by high frequency sound waves: Part II," Proc. Ind. Acad. Sci. Sect. A **2**, 413–420 (1935).
14. C. V. Raman and N. S. N. Nath, "The diffraction of light by high frequency sound waves: Part III," Proc. Ind. Acad. Sci. Sect. A **3**, 75–84 (1936).

Appl. Phys. B 32, 15–20 (1983)

Applied Physics B Photophysics and Laser Chemistry

Diffraction Characteristics of Planar Absorption Gratings

W. E. Baird, M. G. Moharam, and T. K. Gaylord

School of Electrical Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

Received 22 April 1983/Accepted 26 May 1983

Abstract. Planar (co)sinusoidal conductivity (absorption) transmission gratings are analyzed using rigorous coupled-wave theory. The first-order and higher-order diffraction efficiencies are determined over the entire range of possible conductivities and Bragg angles of incidence (or equivalently, grating periods) for H-mode polarization incident plane waves. The maximum possible first diffracted order efficiency is found to be 5.26%. Rigorous results are compared to approximate results from the Raman-Nath theory and the two-wave first-order coupled-wave (Kogelnik) theory. A regime parameter, Q_σ, is defined which delineates the regions of Raman-Nath diffraction behavior ($Q_\sigma < 1$) and the region of two-wave first-order diffraction theory behavior ($Q_\sigma > 1$). Likewise, the angular selectivity characteristics of conductivity gratings are determined from rigorous theory and are compared with corresponding results from approximate theory.

PACS: 42.20, 42.40

Optical diffraction by planar transmission gratings is a subject of fundamental importance in optics. Fields of application include acousto-optics, integrated optics, quantum electronics, holography, and spectroscopy. Grating device functions include laser-beam deflection, modulation, coupling filtering, distributed feedback, distributed Bragg reflection, holographic beam combining, wavelength multiplexing, and wavelength demultiplexing.

A rigorous coupled-wave theory (without approximations) has recently been formulated for lossless dielectric gratings with relative permittivity (dielectric constant) modulation [1]. This analysis has been shown to be general and the approximations used in previous theories have been explicitly quantified [2]. It is the purpose of this paper: 1) to extend the rigorous coupled-wave analysis to (co)sinusoidal conductivity (absorption) gratings, 2) to show that the maximum diffraction efficiency is 5.26% (rather than 3.70% from Kogelnik theory [3] or 4.80% from Raman-Nath theory [4] for these gratings), 3) to define the diffraction regimes and their boundaries for transmission absorption gratings, and 4) to determine rigorously the angular selectivity characteristics of these gratings and compare them to those from approximate theory. To assist in isolating the basic diffraction characteristics from other physical effects, the fundamental case of the same permittivity inside and outside the grating, an unslanted grating (fringes perpendicular to surface), and H-mode polarization (electric field perpendicular to the plane of incidence and perpendicular to the grating vector) is treated.

1. Theory

1.1. Conductivity Grating

The gratings analyzed in this work have a conductivity of the form

$$\sigma(x) = \sigma_0 + \sigma_1 \cos Kx. \tag{1}$$

The grating is unslanted with grating vector **K** (of magnitude $K = 2\pi/\Lambda$, Λ being the grating period) along the x-axis. The planar surfaces of the grating medium are at $z = 0$ and $z = d$. The plane of incidence is the $x-z$ plane and thus all quantities are invariant in the

y-direction. The permittivity ϵ of the grating is constant and equal to the permittivity of the surrounding medium. The permeability μ is that of free space. In terms of these parameters the attenuation factor $\alpha(x)$ is

$$\alpha(x)=\omega(\mu\epsilon)^{1/2}\{\tfrac{1}{2}[[1+(\sigma/\omega\epsilon)^2]^{1/2}-1]\}^{1/2}, \quad (2)$$

where ω is the angular frequency of the incident light wave. The primary quantities of interest here are the diffraction efficiencies of the first-order and higher-order transmitted diffracted waves. In particular, the maximum value of the first-order diffraction efficiency is obtained for the total range of conductivities and grating periods at Bragg incidence.

1.2. Rigorous Coupled-Wave Theory

The rigorous coupled-wave equations for an unslanted (co)sinusoidal conductivity grating for H-mode polarization are

$$\frac{d^2S_i(z)}{dz^2}-\frac{4\pi}{\lambda}\left(\mathrm{j}\frac{\sigma_0}{\omega\epsilon_0}-\varepsilon_0\cos^2\theta'\right)^{1/2}\frac{dS_i(z)}{dz}+\left(\frac{2\pi}{\Lambda}\right)^2 i(m-i)S_i(z)-\mathrm{j}\frac{\pi}{\lambda}\sigma_1\eta_0[S_{i+1}(z)+S_{i-1}(z)]=0, \quad (3)$$

where $S_i(z)$ is the normalized amplitude of the i^{th} space-harmonic field at any point within the modulated region, λ is the free space wavelength of the incident plane wave, ϵ_0 is the permittivity of free space, ε_0 is the relative permittivity (dielectric constant) inside and outside of the grating, θ' is the angle of incidence in the input region,

$$m=2\Lambda\epsilon_0^{1/2}\sin\theta'/\lambda \quad (4)$$

is the Bragg condition for an unslanted absorption grating ($m=1$ for incidence at the first Bragg angle, θ_B), $\eta_0=(\mu_0/\epsilon_0)^{1/2}$ is the characteristic impedance of free space, and μ_0 is the permeability of free space. These rigorous coupled-wave equations can be solved by state-variable methods [5]. Then with the application of electromagnetic boundary conditions (continuity of tangential E and tangential H at $z=0$ and $z=d$), the diffracted fields and thus the diffraction efficiencies can be calculated for any order, reflected or transmitted [1].

1.3. Two-Wave First-Order Theory

In this approximation to the rigorous theory (Kogelnik theory [3]), the only orders retained in the analysis are $i=0$ and $+1$; the second derivatives of field amplitudes are assumed negligible; and the boundary conditions on the two space-harmonic field amplitudes are assumed to be $S_0(0)=1$ and $S_1(0)=0$. The diffraction efficiency for the first-order transmitted wave according to this theory is given by

$$\mathrm{DE}_1=\exp\left(\frac{-\eta_0\sigma_0 d}{\varepsilon_0^{1/2}\cos\theta'}\right)\sinh^2\left(\frac{\eta_0\sigma_1 d}{4\varepsilon_0^{1/2}\cos\theta'}\right) \quad (5)$$

and the zero-order (undiffracted) transmitted efficiency is predicted to be

$$\mathrm{DE}_0=\exp\left(\frac{-\eta_0\sigma_0 d}{\varepsilon_0^{1/2}\cos\theta'}\right)\cosh^2\left(\frac{\eta_0\sigma_1 d}{4\varepsilon_0^{1/2}\cos\theta'}\right). \quad (6)$$

The maximum first-order diffraction efficiency occurs when $\sigma_1=\sigma_0$ and $\eta_0\sigma_0 d/2\varepsilon_0^{1/2}\cos\theta'=\ln 3$. This maximum efficiency has a value of $\mathrm{DE}_{1,\max}=1/27\simeq 3.70\%$. The results of this well-known two-wave, first-order approximation are used as a comparison for the results obtained from rigorous theory.

1.4. Multiwave First-Order Theory Without Dephasing

In this approximation to the rigorous theory (an extension of the Raman-Nath theory of phase gratings [6–8] to absorption gratings [4]), the second derivatives of the space-harmonic field amplitudes are assumed negligible, dephasing from the Bragg condition is ignored, and the boundary conditions on the space-harmonic field amplitudes are assumed to be $S_0(0)=1$ and $S_i(0)=0$ for $i\neq 0$. The diffraction efficiency predicted for any transmitted diffracted order i is given by

$$\mathrm{DE}_i=\exp\left(\frac{-\eta_0\sigma_0 d}{\varepsilon_0^{1/2}\cos\theta'}\right)\mathrm{I}_i^2\left(\frac{\eta_0\sigma_1 d}{2\varepsilon_0^{1/2}\cos\theta'}\right), \quad (7)$$

where I_i is a modified Bessel function of the first kind of integer order i. The quantity i is equal to the diffracted order. The maximum first-order diffraction efficiency occurs when $\sigma_1=\sigma_0$ and $\eta_0\sigma_0 d/2\varepsilon_0^{1/2}\cos\theta'=1.545$ and has a value of $\mathrm{DE}_{1,\max}\simeq 4.80\%$. The results of this multiwave, first-order theory without dephasing are used as a comparison for the results obtained from rigorous theory.

2. Diffraction Characteristics

To determine the diffraction characteristics of planar (co)sinusoidal conductivity gratings, the first-order and higher-order diffraction efficiencies were calculated using the rigorous coupled-wave theory. The maximum first-order transmitted diffraction efficiency was determined for each value of conductivity modulation and Bragg angle of incidence (or equivalently, the grating period). The rigorously-determined diffraction efficiencies were then compared with results from the two-wave first-order coupled-wave (Kogelnik) theory

Table 1. Maximum diffraction efficiencies for sinusoidal conductive gratings. The maximum diffraction efficiencies (given in percent) are shown for each combination of conductivity and Bragg angle. The amplitude of the conductivity modulation in each case is equal to the average conductivity of the grating. The indices of refraction inside and outside of the grating are equal

σ [mho/m]	Angle of incidence (at first Bragg angle)													
			$s^{-1}(1/9)$		$s^{-1}(1/7)$		$s^{-1}(1/5)$		$s^{-1}(1/3)$					
	1°	5°	6.38°	8.21°	10°	11.54°	15°	19.47°	20°	25°	30°	35°	40°	45°
1	3.704	3.700	3.699	3.698	3.698	3.704	3.702	3.704	3.704	3.698	3.703	3.696	3.703	3.704
10	3.760	3.700	3.700	3.691	3.698	3.704	3.702	3.704	3.704	3.698	3.703	3.696	3.703	3.704
10^2	4.687	3.710	3.703	3.700	3.698	3.704	3.702	3.704	3.704	3.698	3.703	3.696	3.703	3.704
10^3	4.800	4.390	4.118	3.836	3.759	3.737	3.714	3.713	3.705	3.698	3.702	3.695	3.701	3.700
5×10^3	4.794	4.775	4.751	4.659	4.498	4.385	4.000	3.857	3.745	3.683	3.675	3.657	3.609	3.575
9,375	4.777	4.797	4.802	4.788	4.743	4.682	4.459	4.251	4.015	3.674	3.607	3.562	3.529	3.454
10^4	4.773	4.791	4.799	4.794	4.754	4.704	4.499	4.319	4.046	3.673	3.596	3.557	3.499	3.432
14,375	4.747	4.775	4.789	4.808	4.807	4.787	4.659	4.627	4.332	3.693	3.515	3.420	3.330	3.230
28,125	4.641	4.666	4.688	4.748	4.787	4.881	4.793	5.126	4.851	3.590	3.275	3.129	3.020	2.906
5×10^4	4.481	4.520	4.552	4.635	4.700	4.837	4.688	5.256	4.906	3.646	3.137	2.927	2.777	2.639
55,937	4.450	4.488	4.512	4.604	4.672	4.812	4.673	5.260	4.911	3.648	3.139	2.900	2.730	2.576
10^5	4.272	4.320	4.362	4.474	4.526	4.717	4.518	4.197	4.849	3.622	3.062	2.721	2.493	2.295
5×10^5	3.441	3.505	3.569	3.744	3.793	4.050	3.736	4.473	4.151	3.141	2.642	2.318	2.114	1.981
10^6	3.053	3.117	3.180	3.353	3.417	3.687	3.408	4.253	3.928	3.001	2.558	2.279	2.102	1.993

and the Raman-Nath theory. The regions of validity of these approximate theories were then delineated. Similarly, the angular selectivity characteristics were calculated using rigorous theory and compared with results from approximate theory.

2.1. Maximum Diffraction Efficiency

The maximum first-order transmitted diffraction efficiencies in percent for a range of Bragg angles of incidence and grating conductivity amplitudes are presented numerically and graphically in Table 1 and Fig. 1, respectively. The conductivity modulation amplitude is always equal to the average conductivity value, as this is necessary for maximum diffraction efficiency. The wavelength of the incident light is 500 nm, and the grating period is varied to keep the angle of incidence always at the first Bragg angle ($m=1$). The relative permittivity (dielectric constant) both inside and outside the grating is the same in order to eliminate the effects due to discontinuities in the average index of refraction. For near-normal incidence and lower values of conductivity, the maximum diffraction efficiency tends to the value of 4.80% predicted by the Raman-Nath multiwave theory, which neglects dephasing. For conditions of near-normal incidence, there are many closely angularly-spaced propagating diffracted orders and dephasing is indeed expected to be of minor importance. For larger Bragg angles of incidence and lower values of conductivity, the maximum diffraction efficiency tends to the value of 3.70% predicted by the Kogelnik two-wave first-order theory. For these larger angles of incidence, the higher-order

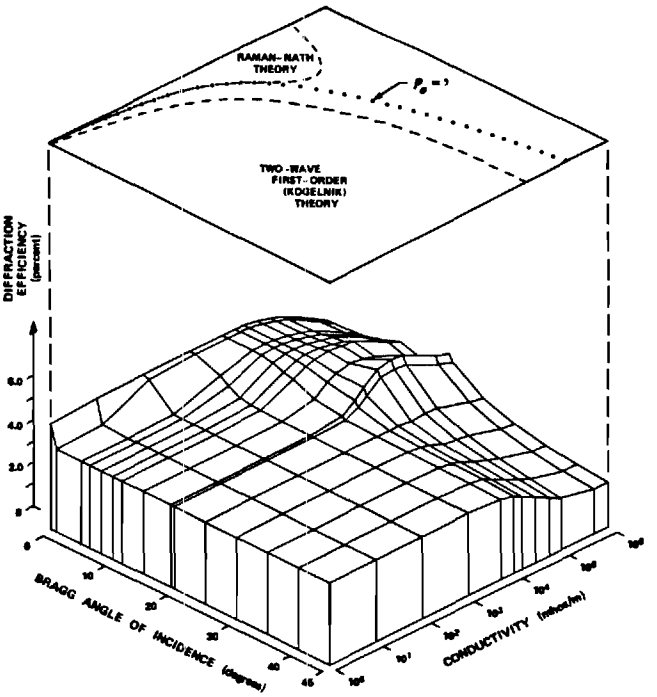


Fig. 1. Maximum diffraction efficiencies for sinusoidal conductive gratings

waves are evanescent and rigorous multiwave theory may be approximated in practice by a two-wave calculation.

A significant structural feature in the resulting maximum diffraction efficiency surface (Fig. 1) occurs for those Bragg angles of incidence at which higher-order diffracted waves are at the transition from propagating to evanescent (cut-off). For example, the angles $\sin^{-1}(1/9)\simeq 6.38°$, $\sin^{-1}(1/7)\simeq 8.21°$, and

Table 2. Example fundamental and higher-order diffraction efficiencies for a Bragg angle incidence of 1.00° for sinusoidal conductivity gratings according to the Raman-Nath, Kogelnik, and rigorous coupled-wave theories. The first case ($\sigma = 1$ mho/m) is in the Bragg regime (Kogelnik theory) and the second case ($\sigma_1 = 10^3$ mho/m) is in the Raman-Nath regime. Other parameters are $\lambda = 0.5$ μm, $\Lambda = 14.325$ μm, $\sigma_1 = \sigma_0$, and thickness chosen to maximize DE_1

Theory	σ_1	ϱ_σ	Diffraction efficiency [%]					
	[mho/m]		DE_0	DE_1	DE_2	DE_3	DE_4	DE_5
Raman-Nath	1	8.12×10^{1}	19.6	4.49	3.05×10^{-1}	9.66×10^{-3}	1.75×10^{-4}	2.05×10^{-6}
Kogelnik	1	8.12×10^{1}	14.9	3.70	–	–	–	–
Rigorous coupled-wave	1	8.12×10^{1}	14.9	3.70	1.38×10^{-4}	5.77×10^{-10}	6.04×10^{-16}	2.28×10^{-22}
Raman-Nath	10^3	8.12×10^{-2}	13.0	4.80	5.98×10^{-1}	3.61×10^{-2}	1.27×10^{-3}	2.93×10^{-5}
Kogelnik	10^3	8.12×10^{-2}	7.84	3.30	–	–	–	–
Rigorous coupled-wave	10^3	8.12×10^{-2}	13.0	4.80	5.98×10^{-1}	3.61×10^{-2}	1.27×10^{-3}	2.85×10^{-5}

$\sin^{-1}(1/5) \simeq 11.54°$ exhibit local diffraction efficiency maxima in the surface and correspond to transitions from 10 to 8 transmitted propagating waves, 8 to 6 waves, and 6 to 4 waves, respectively. For the angle $\sin^{-1}(1/3) \simeq 19.47°$ and a conductivity of 55,937 mho-s/m, the global maximum of 5.260% occurs in the first-order transmitted wave diffraction efficiency. This angle marks the transition from 4 forward-diffracted waves to 2 waves ($i = -1$ and $+2$ become cut-off). Other transitions, of course, occur for specific angles less than $\sin^{-1}(1/9)$. However, the resulting local maxima are masked by the overall Raman-Nath behavior of the surface in that region.

2.2. *Diffraction Regimes*

The regime where the two-wave first-order theory accurately predicts the diffraction characteristics is often referred to as the "Bragg regime". The region where Raman-Nath theory is accurate is called the "Raman-Nath regime". These regions may be distinguished by a regime parameter. The conductivity grating regime parameter ϱ_σ is defined as

$$\varrho_\sigma = \frac{4\pi\lambda}{\eta_0 \sigma_1 \Lambda^2} \tag{8}$$

by analogy to the regime parameter ϱ for phase gratings [9–11] which is

$$\varrho = 2\lambda^2/\varepsilon_1 \Lambda^2, \tag{9}$$

where ε_1 is the amplitude of the relative permittivity modulation of the phase grating. The condition $\varrho_\sigma = 1$ separates the $\sigma - \theta_B$ plane into two regions as shown in Fig. 1. For the region of $\varrho_\sigma > 1$, which includes large Bragg angles of incidence (small grating periods), the two-wave first-order (Kogelnik) result as given by (5) produces accurate results for the fundamental diffracted order ($i = +1$) for conductivities up to about 10^3 mho/m. In the $\varrho_\sigma > 1$ regime, the transmitted power is concentrated primarily in the $i = 0$ and $i = +1$ orders. In fact, the higher-order diffraction efficiencies calculated by rigorous theory were found to obey the condition

$$\sum_{i \neq 0,\, 1} DE_i < 1/\varrho_\sigma^2. \tag{10}$$

That is, the sum of all of the higher-order diffraction efficiencies is less than $1/\varrho_\sigma^2$. This is exactly analogous to the Bragg regime two-wave criterion for phase gratings [11]. Also for $\varrho_\sigma > 1$, the values of the transmitted wave ($i = 0$) efficiency calculated by rigorous theory were compared with the values predicted by (6) from Kogelnik's theory. Good agreement was again found except at high conductivities. An example $\varrho_\sigma > 1$ case showing this agreement is given in Table 2. Since two-wave first-order theory neglects all diffracted orders except the $i = 0$ and $i = +1$ orders, there are no predictions for the higher-order waves using this theory and these are indicated by dashes in Table 2.

For the region of $\varrho_\sigma < 1$, the diffraction efficiencies of all diffracted orders (in addition to the $i = +1$ order) were calculated by rigorous theory and then compared with the values predicted by the Raman-Nath theory, (7). The $\varrho_\sigma < 1$ regime includes near-normal Bragg incidence (large grating periods). In this region the Raman-Nath formula as given by (7) was found to produce accurate results for conductivities up to about 5×10^4 mho/m. This close agreement for the first-order diffracted wave is apparent in Table 1 and Fig. 1. For the zero-order and higher-order diffraction efficiencies, similar good agreement was found. A single typical $\varrho_\sigma < 1$ case showing the agreement with Raman-Nath theory is included in Table 2.

2.3. *Angular Selectivity*

A Bragg condition occurs whenever m in (4) is an integer. Dephasing from the Bragg condition may be produced for a fixed grating by changing the angle of incidence and/or the wavelength. For $m = 1$, it is the

first or fundamental Bragg incidence. In this case, there is efficient power transfer from the incident wave to the $i=+1$ diffracted order. Mathematically, this is due to the factor $(m-i)$ being zero in the rigorous coupled-wave equations, (3). This $S_i(z)$ term in the rigorous coupled-wave equations represents dephasing from the Bragg condition. When it is zero, there is no dephasing and Bragg incidence occurs. The two-wave first-order coupled-wave analysis of Kogelnik retains the effects of dephasing from the Bragg condition. The Raman-Nath theory neglects this term entirely, and any angle of incidence and wavelength is treated as Bragg incidence.

The angular selectivity of a grating is a measure of the sensitivity of the diffraction to changes in the angle of incidence. The angular selectivity, $\Delta\theta$, may be defined as the full angular deviation about the first Bragg angle $(m=1)$ which causes a reduction in the diffraction efficiency to one half the value at the Bragg angle. This angular selectivity may be calculated from rigorous coupled-wave theory or from approximate two-wave first-order coupled-wave theory since these theories include dephasing effects. The angular selectivity is given by

$$\Delta\theta=\theta^{+}-\theta^{-}, \tag{11}$$

where θ^{+} and θ^{-} are the angles of incidence, greater than and less than the Bragg angle, respectively, at which the diffraction efficiency has dropped to one half of the value at the Bragg angle. From two-wave first-order (Kogelnik) theory, these quantities are given by

$$\theta^{\pm}=\sin^{-1}\left\{\frac{\sin\theta_B \pm (\xi\Lambda/\pi d)\left[\cos^2\theta_B-(\xi\Lambda/\pi d)^2\right]^{1/2}}{1+(\xi\Lambda/\pi d)^2}\right\}. \tag{12}$$

The quantity ξ is a dephasing parameter. If $\xi=0$, there is no dephasing and $\theta^{\pm}=\theta_B$ indicating $\Delta\theta=0$ (incidence at Bragg angle). For the maximum efficiency $(DE_{1,max}=1/27)$ in this theory, it is $\xi=0.8952$. The angular selectivity may not be calculated from Raman-Nath theory since that theory does not include any dephasing effects.

A comparison of some angular selectivity results from rigorous theory and from Kogelnik theory are shown in Table 3. In each case the first Bragg angle $\theta_B=30°$, the wavelength $\lambda=500$ nm, and the grating is fully modulated $\sigma_1=\sigma_0$. For each conductivity, the thickness that maximizes the first diffracted order power is used. For relatively thick gratings, the rigorous theory and the Kogelnik theory predict the same angular sensitivities. For high conductivity thin gratings, the angular selectivity, $\Delta\theta$, approaches approximately 80°. However, approximate two-wave first-order (Kogelnik) theory, (12), predicts that the angular selectivity approaches 180° in the limit of increasing conductivity.

Table 3. Angular selectivity for sinusoidal conductive gratings. The full angular deviation about the first Bragg angle, $\Delta\theta$, that causes a reduction in the diffraction efficiency to one half of the value at the Bragg angle is given. Both the approximate value of $\Delta\theta$ from Kogelnik's two-wave first-order coupled-wave theory and the value from the present rigorous theory are shown. In each case $\theta_B=30°$, $\lambda=500$ nm, and the grating is fully modulated. The indices of refraction inside and outside of the grating are equal

σ	d	$\Delta\theta$ [degrees]	
[mho/m]	[mm]	Kogelnik's theory	Rigorous theory
10^{-1}	5.05×10^{1}	5.08×10^{-4}	5.08×10^{-4}
1	5.05	5.08×10^{-3}	5.08×10^{-3}
10	5.05×10^{-1}	5.08×10^{-2}	5.08×10^{-2}
10^{2}	5.05×10^{-2}	5.08×10^{-1}	5.08×10^{-1}
10^{3}	5.05×10^{-3}	5.07	5.08
10^{4}	5.16×10^{-4}	4.69×10^{1}	5.06×10^{1}
10^{5}	8.85×10^{-5}	1.37×10^{2}	7.83×10^{1}
10^{6}	1.07×10^{-5}	1.75×10^{2}	7.79×10^{1}
10^{7}	1.02×10^{-6}	1.79×10^{2}	7.84×10^{1}

3. Summary and Discussion

The rigorous coupled-wave equations for (co)-sinusoidal conductivity (absorption) gratings have been presented. These were then solved subject to the appropriate electromagnetic boundary conditions for the first-order and higher-order transmitted diffraction efficiencies for the entire range of possible conductivities and first Bragg angles of incidence (equivalent to the range of possible grating periods). These results were then compared to results from the Raman-Nath and two-wave first-order (Kogelnik) approximate theories. Example results are shown in Table 2. The global maximum diffraction efficiency for the first-order transmitted diffracted wave was found to be 5.26% rather than 4.80% or 3.70% predicted respectively by the Raman-Nath and Kogelnik approximate theories.

A conductivity grating regime parameter was defined as $\varrho_\sigma=4\pi\lambda/\eta_0\sigma_1\Lambda^2$ by analogy to the phase grating regime parameter [9–11]. The condition $\varrho_\sigma=1$ was shown to delineate Raman-Nath diffraction behavior $(\varrho_\sigma<1)$ and two-wave first-order (Kogelnik) diffraction behavior $(\varrho_\sigma>1)$. For sufficiently high conductivities (about 5×10^{4} mho/m for Raman-Nath theory and about 10^{3} mho/m for Kogelnik theory), it was shown that these approximate theories no longer give accurate results even though the regime parameter condition is met (Fig. 1).

The angular selectivity characteristics of these planar conductivity gratings were analyzed using rigorous coupled-wave theory. Two-wave first-order approximate theory was found to give accurate predictions for conductivities up to about 10^{4} mho/m, but overesti-

mated the angular selectivity for higher conductivities.

H-mode polarization (electric field perpendicular to the plane of incidence and perpendicular to the grating vector) has been analyzed. However, E-mode polarization may be treated in the same manner by starting with the E-mode coupled-wave equations, as shown in [12].

Acknowledgement. This work was sponsored by the National Science Foundation and the Joint Services Electronics Program.

References

1. M.G. Moharam, T.K. Gaylord: J. Opt. Soc. Am. **71**, 811–818 (1981)
2. T.K. Gaylord, M.G. Moharam: Appl. Phys. B**28**, 1–14 (1982)
3. H. Kogelnik: Bell Syst. Tech. J. **48**, 2909–2947 (1969)
4. R. Magnusson, T.K. Gaylord: Opt. Commun. **28**, 1–3 (1979)
5. C.L. Liu, J.W.S. Liu: *Linear Systems Analysis* (McGraw-Hill, New York 1975)
6. C.V. Raman, N.S.N. Nath: Proc. Indian Acad. Sci. A**2**, 406–412 (1935)
7. C.V. Raman, N.S.N. Nath: Proc. Indian Acad. Sci. A**2**, 413–420 (1935)
8. C.V. Raman, N.S.N. Nath: Proc. Indian Acad. Sci. A**3**, 75–84 (1936)
9. N.S.N. Nath: Proc. Indian Acad. Sic. A**8**, 499–503 (1938)
10. M.G. Moharam, L. Young: Appl. Opt. **17**, 1757–1759 (1978)
11. M.G. Moharam, T.K. Gaylord, R. Magnusson: Opt. Commun **32**, 14–18 (1980)
12. M.G. Moharam, T.K. Gaylord: J. Opt. Soc. Am. **73**, 451–455 (1983)

Reprinted from APPLIED OPTICS, Vol. 20, page 3271, October 1, 1981

Thin and thick gratings: terminology clarification

T. K. Gaylord and M. G. Moharam

Georgia Institute of Technology, School of Electrical Engineering, Atlanta, Georgia 30332.
Received 22 May 1981.
0003-6935/81/193271-03$00.50/0.

The terminology *thin* and *thick* gratings is widely used. However, these terms frequently either are not defined, vaguely defined, or defined in an ambiguous way. Interpretations of *thin* and *thick* grating behavior appear in the literature dating back to the 1930's. These interpretations with their various degrees of preciseness and accuracy have been carried forward in parallel in many cases. The terminology of *thin* and *thick* gratings is often confusing to workers in fields that use planar gratings (such as acoustooptics, holography, integrated optics, and spectroscopy). The purpose of this short paper is to clarify the possible practical explicit definitions of *thin* and *thick* gratings. This is done in terms of the diffraction regime characteristics and angular and wavelength selectivity characteristics of the grating. For brevity, only the common case of planar gratings with grating fringes perpendicular to the surfaces is discussed. However, planar absorption gratings and slanted fringe gratings may be similarly analyzed.

For a plane wave incident upon a planar grating with (co)-sinusoidal permittivity fringes normal to the surface, the fields inside the grating may be completely described by the rigorous coupled-wave equations.[1] These equations are obtained by substituting the periodic relative permittivity, the plane wave field expansion, and the Floquet theorem into the wave equation. For H-mode polarization the result is

$$\frac{1}{2\pi^2}\frac{d^2S_i(z)}{dz^2} - j\frac{2(\epsilon_0)^{1/2}\cos\theta}{\pi\lambda}\frac{dS_i(z)}{dz} + \frac{2i(m-i)}{\Lambda^2}S_i(z) + \frac{\epsilon_1}{\lambda^2}[S_{i+1}(z) + S_{i-1}(z)] = 0, \quad (1)$$

where $S_i(z)$ are the fields inside the grating, i is the (integer) order of diffraction, $(-\infty < i < +\infty)$, ϵ_0 is the average relative permittivity, ϵ_1 is the amplitude of the (co)sinusoidal relative permittivity, λ is the free space wavelength, θ is the angle of refraction of the incident wave, Λ is the grating period, and m is given by

$$m\lambda/(\epsilon_0)^{1/2} = 2\Lambda\sin\theta. \quad (2)$$

This is a Bragg condition when m is an integer. Equation (1) is derived without approximation.

The Raman-Nath diffraction regime[2] may be obtained from Eq. (1) by neglecting the d^2S_i/dz^2 and S_i terms. The Bragg diffraction regime may be obtained from Eq. (1) by neglecting the d^2S_i/dz^2 term and considering only $i = 0,1$. This corresponds to the Kogelnik two-wave coupled-wave theory.[3] In any case, the diffraction efficiency is given by $\eta_i = S_iS_i^*$ for unslanted phase gratings.

A *thin* grating may be described as a grating that produces Raman-Nath regime diffraction.[2] In this case, the multiple grating diffracted-orders ideally have diffraction efficiencies η_i given by

$$\eta_i = J_i^2(2\gamma), \quad (3)$$

where i is the integer representing the diffracted-order, J_i is an integer-order ordinary Bessel function of the first kind, and γ is the grating strength parameter given by $\gamma = \pi\epsilon_1 d/2\lambda(\epsilon_0)^{1/2}\cos\theta$, and d is the grating thickness. The occurrence (or lack of occurrence) of Raman-Nath regime diffraction as given by Eq. (3) may be determined using a number of different criteria[4] depending on the application of the grating. The practical criteria and their interpretations are: (1) *Zero-Order Beam Criterion*—the transmitted (zero-order) diffraction efficiency is predicted by Eq. (3) to within some specified limit; (2) *First-Order Beam Criterion*—the fundamental (first-order) diffraction efficiency is predicted by Eq. (3) to within some specified limit; and (3) *Composite Criterion*—all diffracted-orders simultaneously have diffraction efficiencies predicted by Eq. (3) to within some specified limit. Each of these criteria is evaluated in Ref. 4. Those evaluations showed that each of these criteria is met to within 1% diffraction efficiency when the condition

$$Q'\gamma \le 1 \quad (4)$$

is satisfied. Q' is a grating parameter given by $Q' = Q/\cos\theta = 2\pi\lambda d/(\epsilon_0)^{1/2}\Lambda^2\cos\theta$. This condition was originally used by Extermann and Wannier[5] and later in the form of $Q'\gamma \le \pi^2/8$ by Willard.[6] From Eq. (4) it is apparent that Raman-Nath regime diffraction behavior will be observed for any value of γ (proportional to grating modulation ϵ_1) if Q' is sufficiently small. This has led to the incomplete popular condition of $Q' < 1$ for describing *thin* gratings.[7] The first-order diffraction efficiency $J_1^2(2\gamma)$ for ideal Raman-Nath regime diffraction is shown in Fig. 1.

A *thick* grating (or *volume* grating) may be described as a grating that produces Bragg regime (or two-wave regime) diffraction. This is described by the two-wave coupled-wave

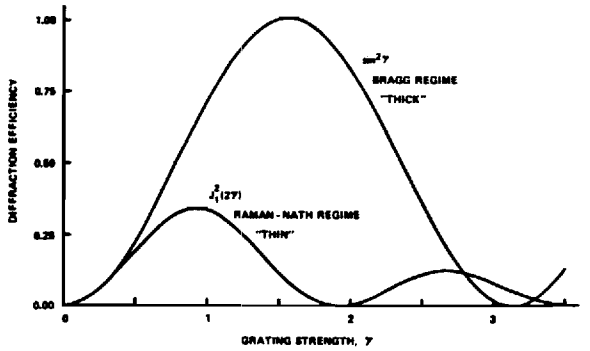


Fig. 1. Ideal first-order diffraction efficiencies for Raman-Nath (*thin* grating) regime [$\eta_1 = J_1^2(2\gamma)$] and for Bragg (*thick* grating) regime ($\eta_1 = \sin^2\gamma$).

theory of Kogelnik.[3] In this regime, the single fundamental diffracted-order ideally has a diffraction efficiency given by

$$\eta_1 = \sin^2\gamma. \tag{5}$$

Correspondingly, the transmitted (zero-order) wave has a diffraction efficiency given by $\cos^2\gamma$. For E-mode polarization the γ in Eq. (5) is $\gamma = \pi\epsilon_1 d \cos 2\theta/2\lambda(\epsilon_0)^{1/2}\cos\theta$. The occurrence (or lack of occurrence) of Bragg regime diffraction as given by Eq. (5) may be determined using a number of different criteria[8] depending on the application. The practical criteria include the *Zero-Order Beam Criterion* and *First-Order Beam Criterion* defined as before except that the deviation is with respect to Eq. (5). The other criteria and their interpretations are: *Two-Wave Criterion*—the sum of the diffraction efficiency associated with all higher-order waves is less than some specified limit; and *Composite Criterion*—all three criteria are met to within some specified limit. Each of these criteria is evaluated in Ref. 8. Those evaluations showed that each of these criteria is met to within 1% diffraction efficiency when the condition

$$\rho \equiv Q'/2\gamma \geq 10 \tag{6}$$

is satisfied. The parameter ρ was first used by Nath.[9] It is now clear that this parameter determines the Bragg regime diffraction boundary. Since $\rho = 2\lambda^2/\Lambda^2\epsilon_1$ for H mode and $\rho = 2\lambda^2/\Lambda^2\epsilon_1\cos 2\theta$ for E mode, it is particularly notable that this diffraction-regime-based definition of a *thick* grating is independent of grating thickness. If, in addition to the above criteria, it is also required that $\eta_1 > \eta_{-1}$ (for example, in holography, to make the power in the conjugate image beam be small compared with that in the primary diffracted beam), the condition $Q' > 1$ is also needed in addition to (6).[10,11] From Eq. (6) it is apparent that Bragg regime diffraction behavior will occur for any value of γ (or ϵ_1) if Q' is sufficiently large. Thus, the incomplete condition $Q' > 1$ is often used alone to describe *thick* gratings.[3,7]

For exact Bragg regime diffraction, the diffraction efficiency is given by $\sin^2\gamma$ and this is also shown in Fig. 1. For small values of the grating strength γ, the diffraction efficiencies of the fundamental diffracted-order converge to the same diffraction efficiency $\eta_1 = \gamma^2$ for both *thin* and *thick* gratings. When both conditions (4) and (6) are simultaneously satisfied the distinction between *thin* and *thick* is lost when using diffraction-regime-based definitions.

When neither condition (4) nor (6) is satisfied, the grating will be neither *thin* nor *thick* according to diffraction-regime-based definitions. The first-order diffraction efficiency will not be accurately predicted by either Eqs. (3) or (5). In general, in the intermediate regime, a rigorous (without approximations) planar grating diffraction theory (e.g., Refs. 1, 12, and 13) is required.

A *thin* grating may be alternatively described as a grating exhibiting relatively little angular and wavelength selectivity. As the incident wave is dephased (either in angle incidence or in wavelength) from the Bragg condition, the diffraction efficiency decreases. The angular range or wavelength range for which the diffraction efficiency decreases to half of its on-Bragg-angle value is determined by the thickness of the grating d expressed as a number of grating periods Λ. For a *thin* grating this number may be reasonably chosen to be

$$d/\Lambda < 10. \tag{7}$$

The region of *thin* grating behavior according to the angular-and-wavelength-selectivity-based definition (7) is depicted in Fig. 2. Gratings having angular and wavelength selectivities with FWHM wider than that for $d/\Lambda = 10$ may be considered to be *thin* gratings. This definition does not accurately predict the diffraction regime. It has the desirable feature that the governing parameter (d/Λ) is directly proportional to the grating thickness, and thus *thin* and *thick* have direct physical interpretations.

A *thick* grating may conversely be described as a grating exhibiting strong angular and wavelength selectivity. A relatively small change in the angle of incidence from the Bragg angle or a relatively small change in the wavelength at the Bragg angle produces significant dephasing and the diffraction efficiency decreases correspondingly. *Thick* grating behavior may be considered to occur when

$$d/\Lambda > 10. \tag{8}$$

This is the angular-and-wavelength-selectivity-based definition of a *thick* grating. The region of this behavior is also shown in Fig. 2. Unlike with the diffraction-regime-based definitions ($\gamma Q' \leq 1$ and $Q'/2\gamma \geq 10$), there is a single simple boundary ($d/\Lambda = 10$) between *thin* and *thick* grating diffraction.

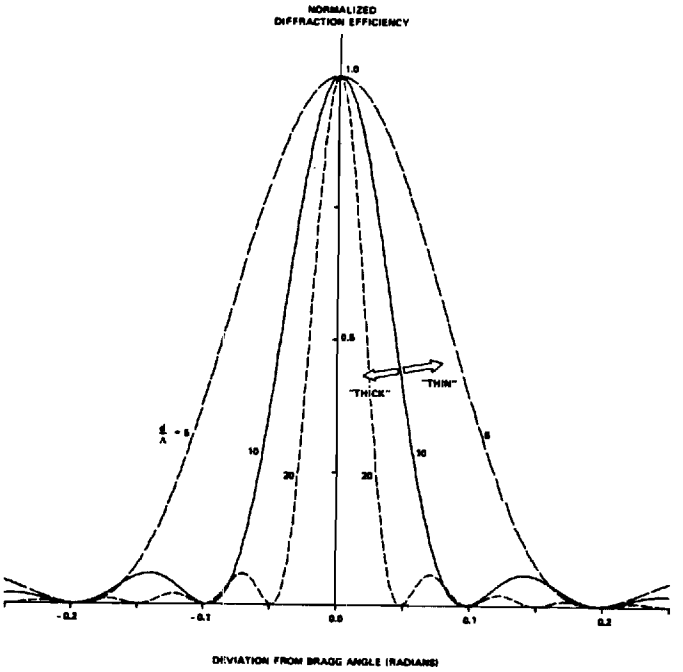


Fig. 2. Typical angular selectivity plots (normalized first-order diffraction efficiency vs angular deviation from the Bragg angle) for various values of d/Λ.

The terms *thin* and *thick* gratings may be ambiguous. However, distinct meaningful definitions are possible based on either the diffraction regime or on the angular and wavelength selectivity. These definitions may be concisely summarized as follows:

(1) If *thin* grating is intended to mean Raman-Nath regime diffraction, then the required condition is $Q'\gamma \leq 1$.

(2) If *thin* grating is intended to mean broad angular and wavelength selectivity, then the required condition is $d/\Lambda < 10$.

(3) If *thick* grating is intended to mean Bragg regime diffraction, then the required condition is $\rho \geq 10$. (If $\eta_1 > \eta_{-1}$ is also included in the definition of Bragg regime, $Q' > 1$ is required in addition.)

(4) If *thick* grating is intended to mean narrow angular and wavelength selectivity, then the required condition is $d/\Lambda > 10$.

This work was sponsored by the National Science Foundation and the Joint Services Electronics Program.

References

1. M. G. Moharam and T. K. Gaylord, J. Opt. Soc. Am. **71**, 811 (1981).
2. C. V. Raman and N. S. N. Nath, Proc. Indian Acad. Sci. **2**, 406, 413 (1935); **3**, 75, 119 (1936).
3. H. Kogelnik, Bell Syst. Tech. J. **48**, 2909 (1969).
4. M. G. Moharam, T. K. Gaylord, and R. Magnusson, Opt. Commun. **32**, 19 (1980).
5. R. Extermann and G. Wannier, Helv. Phys. Acta **9**, 520 (1936).
6. G. W. Willard. J. Acoust. Soc. Am. **21**, 101 (1949).
7. W. R. Klein and B. D. Cook, IEEE Trans. Sonics Ultrason. **SU-14**, 123 (1967).
8. M. G. Moharam, T. K. Gaylord, and R. Magnusson, Opt. Commun. **32**, 14 (1980).
9. N. S. N. Nath, Proc. Indian Acad. Sci. **8**, 499 (1938).
10. P. Phariseau, Proc. Indian Acad. Sci. Sect. A **44**, 165 (1965).
11. B. Benlarbi, D. J. Cooke, and L. Solymar, Opt. Acta. **27**, 885 (1980).
12. C. B. Burckhardt, J. Opt. Soc. Am. **56**, 1502 (1966).
13. R. S. Chu and T. Tamir, IEEE Trans. Microwave Theory Tech. **MTT-18**, 486 (1970).

Three-dimensional vector coupled-wave analysis of planar-grating diffraction

M. G. Moharam and T. K. Gaylord

School of Electrical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

Received March 24, 1983

Diffraction by an arbitrarily oriented planar grating with slanted fringes is analyzed using rigorous three-dimensional vector coupled-wave analysis. The method applies to any sinusoidal or nonsinusoidal amplitude and/or phase grating, any plane-wave angle of incidence, and any linear polarization. In the resulting (conical) diffraction, it is shown that coupling exists between all space-harmonic vector fields inside the grating (corresponding to diffracted orders outside the grating). Therefore the TE and TM components of an incident wave are each coupled to all the TE and TM components of all the forward- and backward-diffracted waves. For a general Bragg angle of incidence, it is shown that the diffraction efficiency can approach 100% for a lossless grating if either the incident electric field or the magnetic field is perpendicular to the grating vector. Maximum coupling between incident and diffracted waves is shown to occur when the incident electric field is perpendicular to the grating vector. In general, the diffracted waves are shown to be elliptically polarized. The three-dimensional vector coupled-wave analysis presented is shown to reduce to ordinary rigorous coupled-wave theory when the grating vector lies in the plane of incidence.

INTRODUCTION

Planar amplitude and phase gratings are of wide interest owing to their many applications in quantum electronics, integrated optics, acousto-optics, spectroscopy, and holography. Example grating devices include distributed-feedback lasers, beam deflectors, beam modulators, waveguide couplers, spectral filters, wavelength multiplexers and demultiplexers, and holographic beam combiners.

The most common methods of analyzing planar grating diffraction are the coupled-wave approach[1-9] and the modal approach.[10-19] These investigations were restricted to the case of a grating vector lying in the plane of incidence (the plane defined by the wave normal and the boundary normal). In this situation, the TE (electric field perpendicular to the plane of incidence) and the TM (magnetic field perpendicular to plane of incidence) components of the input plane wave are completely decoupled and may be treated separately. In this special case, (1) if the incident plane wave has TE polarization, the grating-diffraction problem is described as having H-mode polarization since the magnetic field lies in the plane of the wave normal and the grating vector (the electric field is perpendicular to the grating vector) and (2) if the incident plane wave has TM polarization, the grating-diffraction problem is described as having E-mode polarization since the electric field lies in the plane of the wave normal and the grating vector (the magnetic field is perpendicular to the grating vector). However, in the general case, as treated in this paper, the grating vector may have any arbitrary orientation with respect to the plane of incidence. In this situation, the TE and TM components of the input plane wave are coupled inside the grating and may not be treated separately. In this general case the grating-diffraction problem may *not* be decomposed into separate TE- and TM-polarization problems, as is usually done.

The description of grating diffraction as a direct solution of Maxwell's equations has been considered by Neviére *et al.*,[20-22] by Chang *et al.*,[23] and by Knop.[24] Two first-order Maxwell equations were solved directly rather than by the usual procedure of solving a single second-order wave equation. In these analyses, the grating vector was restricted to lie in the plane of incidence. The general three-dimensional case in which the grating vector is not in the plane of incidence is sometimes called conical diffraction (for reasons described below). This case has been discussed by Maystre[25] and has been treated using a Green function approach by Chuang and Kong.[26] The present work combines the direct solution of Maxwell's equation and the general three-dimensional diffraction geometry.

THEORY

The general three-dimensional grating-diffraction problem is depicted in Fig. 1. A linearly polarized electromagnetic wave is obliquely incident at an arbitrary angle α on a slanted-fringe nonsinusoidal mixed amplitude and phase planar grating of slant angle ϕ bounded by two different homogeneous media. The planar grating has an arbitrary direction of periodicity (direction of grating vector **K**). There are four fundamental directions that specify this grating-diffraction geometry: (1) the wave-normal direction of the incident wave, (2) the electric-field direction (polarization) of the incident wave, (3) the normal to the planar grating boundaries, and (4) the grating vector. In the analysis presented here, without any loss of generality, the following geometry is used: (1) the boundary normals are in the z direction, (2) the grating vector is in the x–z plane, and (3) the plane of incidence makes an angle δ with respect to the x axis.

The modulated region ($0 < z < d$) contains a mixed amplitude and phase grating. The grating may be characterized

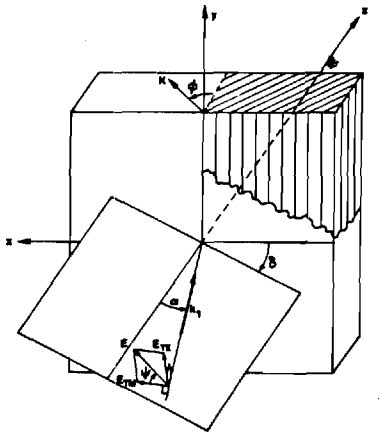


Fig. 1. Geometry of a slanted-fringe planar grating with a plane wave of wave vector $\mathbf{k}_1$ incident at an arbitrary angle and with arbitrary linear polarization.

by a periodic complex relative permittivity (dielectric constant) expandable in a Fourier series as

$$\varepsilon(x, z) = \sum_h \hat{\varepsilon}_h \exp(jh\mathbf{K} \cdot \mathbf{r}). \tag{1}$$

The quantity $\hat{\varepsilon}_h$ is the hth Fourier component of the complex relative permittivity and is given by $\hat{\varepsilon}_h = \varepsilon_h - j\sigma_h/\omega\epsilon_0$, where ε_h and σ_h are the hth Fourier components of the real dielectric constant and the conductivity, respectively. The quantity ω is the angular frequency of the incident radiation, and ϵ_0 is the permittivity of free space. The grating vector is given by

$$\begin{aligned}\mathbf{K} &= K_x\hat{x} + K_z\hat{z} \\ &= K \sin\phi\hat{x} + K\cos\phi\hat{z},\end{aligned} \tag{2}$$

where $K = 2\pi/\Lambda$, Λ is the grating period, and ϕ is the slant angle (angle between the z axis and the grating vector).

In region 1, the incident normalized electric-field vector is

$$\mathbf{E}_{\text{inc}} = \hat{u}\exp(-j\mathbf{k}_1 \cdot \mathbf{r}), \tag{3}$$

with

$$\mathbf{k}_1 = k_1(\sin\alpha\cos\delta\hat{x} + \sin\alpha\sin\delta\hat{y} + \cos\alpha\hat{z}), \tag{4}$$

where α is the angle of incidence (the angle between the wave normal $\mathbf{k}_1$ and the z axis), δ is the angle between the plane of incidence and the x axis, $k_1 = k\varepsilon_I^{1/2}$, ε_I is the relative permittivity in region 1, $k = 2\pi/\lambda$, λ is the free-space wavelength, and $\hat{u}$ is the polarization unit vector given by

$$\begin{aligned}\hat{u} &= u_x\hat{x} + u_y\hat{y} + u_z\hat{z} \\ &= (\cos\psi\cos\alpha\cos\delta - \sin\psi\sin\delta)\hat{x} \\ &\quad + (\cos\psi\cos\alpha\sin\delta \\ &\quad + \sin\psi\cos\delta)\hat{y} \\ &\quad - \cos\psi\sin\alpha\hat{z},\end{aligned} \tag{5}$$

where ψ is the angle between the polarization vector and the plane of incidence. For $\psi = 0°$ and $\psi = 90°$, the magnetic field and the electric field, respectively, are perpendicular to the plane of incidence. The general approach to solve the exact electromagnetic boundary value problem associated with the diffraction grating is to find solutions that satisfy Maxwell's equations (or the corresponding wave equations) in each of the three regions and then to match the tangential electric and magnetic fields at the two boundaries ($z = 0$ and $z = d$). In the general three-dimensional problem, the polarization cannot be decomposed into H-mode and E-mode components with each of these treated separately and then the results combined to obtain the total diffracted field. All the field components are coupled to one another, and solutions for all the electric-field and magnetic-field components have to be obtained simultaneously. The normalized total vector electric field in region 1 ($z < 0$) and in region 3 ($z > d$) may be expressed as

$$\mathbf{E}_1 = \mathbf{E}_{\text{inc}} + \sum_i \mathbf{R}_i \exp(-j\mathbf{k}_{1i} \cdot \mathbf{r}), \tag{6}$$

$$\mathbf{E}_3 = \sum_i \mathbf{T}_i \exp[-j\mathbf{k}_{3i} \cdot (\mathbf{r} - \mathbf{d})], \tag{7}$$

where $\mathbf{R}_i$ is the normalized vector electric field of the ith backward-diffracted (reflected) wave in region 1 with wave vector $\mathbf{k}_{1i}$ and $\mathbf{T}_i$ is the normalized vector electric field of the ith forward-diffracted (transmitted) wave in region 3 with wave vector $\mathbf{k}_{3i}$. Note that, for plane waves, $\mathbf{k}_{1i} \cdot \mathbf{R}_i = 0 = \mathbf{k}_{3i} \cdot \mathbf{T}_i$. Phase matching and the Floquet theorem require that

$$\begin{aligned}\mathbf{k}_{li} &= [(\mathbf{k}_1 - i\mathbf{K}) \cdot \hat{x}]\hat{x} + [(\mathbf{k}_1 - i\mathbf{K}) \cdot \hat{y}]\hat{y} + k_{zli}\hat{z} \\ &= k_{xi}\hat{x} + k_y\hat{y} + k_{zli}\hat{z},\end{aligned} \tag{8}$$

where

$$k_{xi} = k_1 \sin\alpha\cos\delta - iK\sin\phi, \tag{9}$$

$$k_y = k_1 \sin\alpha\sin\delta, \tag{10}$$

$$k_{zli} = (k_l^2 - k_{xi}^2 - k_y^2)^{1/2} \tag{11}$$

for $l = 1, 3$ (the region index), $k_3 = k\varepsilon_{III}^{1/2}$, and ε_{III} is the average relative pemittivity in region 3. The z component of the wave vector, k_{z3i}, is either positive real (a propagating wave) or negative imaginary (an evanescent wave). Likewise, for region 1, k_{z1i} is either negative real (propagating wave) or positive imaginary (evanescent wave).

The geometrical parameters associated with the diffracted waves are shown in Fig. 2. Region 3 is shown in Fig. 2, but the parameters shown also apply to the diffracted waves in region 1. The angle of diffraction for the ith propagating order is given by

$$\tan\beta_{li} = (k_{xi}^2 + k_y^2)^{1/2}/k_{zli}, \tag{12}$$

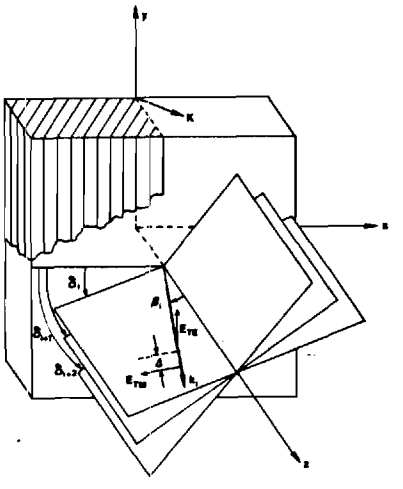


Fig. 2. Geometry associated with the ith forward-diffracted wave.

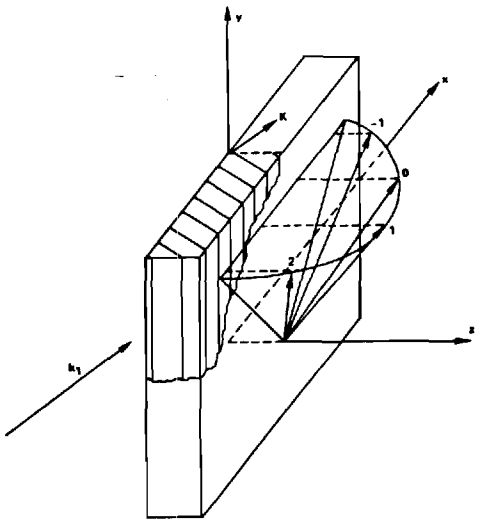


Fig. 3. Geometry of forward-diffracted wave vectors showing conical nature of diffraction. All forward-diffracted waves ($i = -1$ to $i = +2$) have wave vectors that are equal in magnitude and have the same y component.

and the angle of inclination of the output plane is given by

$$\tan \delta_i = k_y/k_{xi}. \tag{13}$$

The plane of diffraction, in general, is different for each diffracted order, as shown in Fig. 2. In the limiting case when the plane of incidence is the x–z plane, $k_y = 0$ and all diffracted orders lie in the same plane (the plane of incidence). However, if the plane of incidence does not contain the grating vector, the k_y is a nonzero constant. The wave vectors of all diffracted orders (forward and backward) have the same y component (perpendicular to the grating vector). This is more clearly shown in Fig. 3, in which the forward-diffracted waves in region 3 are depicted. The magnitude of the wave vectors for all diffracted waves is k_l. This, together with a constant value for k_y for all diffracted orders, means that the wave vectors lie on the surface of a cone (with the cone axis in the y direction); hence the terminology "conical diffraction" for this general three-dimensional geometry. (In general, the cone axis is in the $\hat{n} \times \mathbf{K}$ direction, where $\hat{n}$ is the normal to the boundary.)

The magnetic-field vector in regions 1 and 3 can be obtained by using the Maxwell equation

$$\mathbf{H} = (j/\omega\mu_0)\nabla \times \mathbf{E}, \tag{14}$$

where μ_0 is the permeability of free space, which is the assumed permeability in all regions.

In the modulated region ($0 < z < d$), the electric and magnetic fields may be expressed as Fourier expansions in terms of the space-harmonic fields as

$$\mathbf{E}_2 = \sum_i [S_{xi}(z)\hat{x} + S_{yi}(z)\hat{y} + S_{zi}(z)\hat{z}]\exp(-j\boldsymbol{\sigma}_i \cdot \mathbf{r}), \tag{15}$$

$$\mathbf{H}_2 = (\epsilon_0/\mu_0)^{1/2} \sum_i [U_{xi}(z)\hat{x} + U_{yi}(z)\hat{y} + U_{zi}(z)\hat{z}]\exp(-j\boldsymbol{\sigma}_i \cdot \mathbf{r}), \tag{16}$$

where

$$\boldsymbol{\sigma}_i = k_{xi}\hat{x} + k_y\hat{y} - iK_z\hat{z}. \tag{17}$$

The x and y components of $\boldsymbol{\sigma}_i$ are determined by the Floquet theorem and the phase-matching condition; the z component is arbitrary and can be included in the $\mathbf{S}_i(z)$ and $\mathbf{U}_i(z)$ functions. The z component, however, is chosen in Eq. (17) so that the differential coupled-wave equations to be derived later will have constant coefficients and will thus be directly solvable by the state-variable method. $\mathbf{S}_i(z)$ and $\mathbf{U}_i(z)$ are the normalized amplitudes of the ith space-harmonic vector electric and vector magnetic fields such that $\mathbf{E}_2$ and $\mathbf{H}_2$ satisfy Maxwell's equations (or the appropriate wave equations derived from Maxwell's equations) in the grating region.

For the case when $\phi \equiv 0$, the grating vector is normal to the boundary. Such gratings are called pure reflection gratings. The grating fringes (surfaces of constant ε) in this case are parallel to the grating boundaries ($z = 0, d$). Since the permittivity is no longer periodic along the boundary, the field inside the grating can no longer be expanded in terms of space-harmonic components. However, this pure-reflection-grating case can be simply analyzed without approximation by using a rigorous chain-matrix method of analysis.[27]

METHOD OF SOLUTION

In the general three-dimensional vectorial problem under consideration, all the electric and magnetic space-harmonic fields are coupled to one another. Therefore, rather than attempting to construct and solve two complicated vector wave equations, it is more convenient and straightforward to solve Maxwell's equations

$$\nabla \times \mathbf{E}_2 = -j\omega\mu_0\mathbf{H}_2, \tag{18}$$

$$\nabla \times \mathbf{H}_2 = j\omega\epsilon_0\varepsilon(x, z)\mathbf{E}_2 \tag{19}$$

directly. Substituting Eqs. (15) and (16) into these two equations, and eliminating the components of $\mathbf{E}_2$ and $\mathbf{H}_2$ normal to the boundary, results in a set of four first-order coupled-wave equations:

$$\frac{dS_{xi}(z)}{dz} = -j\left\{iK_zS_{xi}(z) + (k_{xi}/k)\sum_p a_{i-p}[k_yU_{xp}(z) - k_{xp}U_{yp}(z)] + kU_{yi}(z)\right\}, \tag{20}$$

$$\frac{dS_{yi}(z)}{dz} = -j\left\{iK_zS_{yi}(z) - kU_{xi}(z) + (k_y/k)\sum_p a_{i-p}[k_yU_{xp}(z) - k_{xp}U_{yp}(z)]\right\}, \tag{21}$$

$$\frac{dU_{xi}(z)}{dz} = j\left\{(k_{xi}/k)[k_yS_{xi}(z) - k_{xi}S_{yi}(z)] + k\sum_p \hat{\varepsilon}_{i-p}S_{yp}(z) - iK_zU_{xi}(z)\right\}, \tag{22}$$

$$\frac{dU_{yi}(z)}{dz} = -j\left\{k\sum_p \hat{\varepsilon}_{i-p}S_{xp}(z) - (k_y/k)[k_yS_{xi}(z) - k_{xi}S_{yi}(z)] + iK_zU_{yi}(z)\right\}, \tag{23}$$

where $p = i - h$ and a_h is the hth coefficient of the Fourier expansion of $\varepsilon^{-1}(x, z)$ in the form

$$\varepsilon^{-1}(x, z) = \sum_h a_h \exp(jh\mathbf{K} \cdot \mathbf{r}). \tag{24}$$

Note that, when the grating vector is in the plane of incidence, $k_y = 0$ and the coupled-wave equations [Eqs. (20)–(23)] are reduced to two sets of coupled-wave equations; the first set [Eqs. (20) and (23)] gives the solution for the E-mode-polarization case, and the second set [Eqs. (21) and (22)] gives the solution for H-mode polarization.

The coupled-wave equations [Eqs. (20)–(23)] may be written in a matrix form as

$$\begin{bmatrix} \dot{S}_{xi} \\ \dot{S}_{yi} \\ \dot{U}_{xi} \\ \dot{U}_{yi} \end{bmatrix} = \left[\begin{array}{c|c|c|c} a & 0 & c & d \\ \hline 0 & f & g & h \\ \hline i & j & k & 0 \\ \hline m & n & 0 & p \end{array}\right] \begin{bmatrix} S_{xi} \\ S_{yi} \\ U_{xi} \\ U_{yi} \end{bmatrix} \tag{25}$$

or in compact form as

$$\dot{\mathbf{V}} = \mathbf{A}\mathbf{V}, \tag{26}$$

where $\mathbf{V}$ is a vector composed of S_{xi}, S_{yi}, U_{xi}, and U_{yi}. The coefficient matrix $\mathbf{A}$ is the system matrix composed of the 16 submatrices in Eq. (25) that are in turn specified by the 4 sets of coupled-wave equations.

Equation (26) may be solved using a state-variable method (described in detail in previous publications[8,28]) by calculating the eigenvalues and eigenvectors associated with the matrix $\mathbf{A}$. The solutions of the coupled-wave equations using the state-variable method may be expressed as

$$S_{xi}(z) = \sum_m C_m w_{1,im} \exp(\lambda_m z), \tag{27}$$

$$S_{yi}(z) = \sum_m C_m w_{2,im} \exp(\lambda_m z), \tag{28}$$

$$U_{xi}(z) = \sum_m C_m w_{3,im} \exp(\lambda_m z), \tag{29}$$

$$U_{yi}(z) = \sum_m C_m w_{4,im} \exp(\lambda_m z), \tag{30}$$

where C_m's are the unknown constants to be determined from boundary conditions, λ_m's are the eigenvalues of the matrix $\mathbf{A}$, and $w_{q,im}$'s are the elements of the eigenvector matrices corresponding to a given value of i (space-harmonic field inside the grating or diffracted order outside the grating). The eigenvalues and eigenvectors are typically calculated by using a computer library program.[29] Note that, if n space harmonics (values of i) are retained in the analysis, the matrix $\mathbf{A}$ will be $4n \times 4n$ and the vector $\mathbf{V}$ will be of length $4n$. This system of equations produces $4n$ eigenvalues and $4n$ values of the unknown constants C_m, and each of the four eigenvector submatrices ($q = 1, 4$) $w_{q,im}$ will be an $n \times 4n$ matrix (n values of i and $4n$ values of m).

The amplitudes of the diffracted fields $\mathbf{R}_i$ and $\mathbf{T}_i$ (together with C_m) are calculated by matching the tangential electric and magnetic fields at the two boundaries. At $z = 0$,

$$u_x \delta_{i0} + R_{xi} = S_{xi}(0), \tag{31}$$

$$u_y \delta_{i0} + R_{yi} = S_{yi}(0), \tag{32}$$

$$\delta_{i0}(k_y u_z - k_1 \cos\alpha u_y) - k_{z1i} R_{yi} + k_y R_{zi} = k U_{xi}(0), \tag{33}$$

$$\delta_{i0}(k_1 \cos\alpha u_x - k_{x0} u_z) + k_{z1i} R_{xi} - k_{xi} R_{zi} = k U_{yi}(0). \tag{34}$$

At $z = d$,

$$T_{xi} = S_{xi}(d)\exp(jiK_z d), \tag{35}$$

$$T_{yi} = S_{yi}(d)\exp(jiK_z d), \tag{36}$$

$$-k_{z3i} T_{yi} + k_y T_{zi} = k U_{xi}(d)\exp(jiK_z d), \tag{37}$$

$$k_{z3i} T_{xi} - k_{xi} T_{zi} = k U_{yi}(d)\exp(jiK_z d). \tag{38}$$

Note that, since $\mathbf{k}_{1i} \cdot \mathbf{R}_i = 0$ and $\mathbf{k}_{3i} \cdot \mathbf{T}_i = 0$, then

$$k_{xi} R_{xi} + k_y R_{yi} + k_{z1i} R_{zi} = 0, \tag{39}$$

$$k_{xi} T_{xi} + k_y T_{yi} + k_{z3i} T_{zi} = 0. \tag{40}$$

The system of simultaneous linear equations given by Eqs. (31)–(40) may be solved for $\mathbf{R}_i$ and $\mathbf{T}_i$ (by using a technique such as Gauss elimination with the maximum pivot strategy[30]). Note that the number of equations is exactly equal to the number of unknowns. For example, if n waves (values of i) are retained in the analysis, there will be $4n$ values of C_m and $3n$ components of each of $\mathbf{R}_i$ and $\mathbf{T}_i$. Thus the total number of unknowns is $10n$, which is exactly the number of equations given by Eqs. (31)–(40). The computational time and storage requirements may be reduced appreciably by eliminating $\mathbf{R}_i$ and $\mathbf{T}_i$ from Eqs. (31)–(38) and solving for the C_m's and then calculating $\mathbf{R}_i$ and $\mathbf{T}_i$.

The diffraction efficiency is defined as the ratio of the component of the real power carried by the diffracted wave normal to the boundary (z component) to the corresponding component of the real power associated with the incident wave. That is,

$$\mathrm{DE}_{1i} = -\mathrm{Re}(k_{z1i}/k_1 \cos)|\mathbf{R}_i|^2, \tag{41}$$

$$\mathrm{DE}_{3i} = \mathrm{Re}(k_{z3i}/k_1 \cos)|\mathbf{T}_i|^2, \tag{42}$$

where DE_{1i} and DE_{3i} are the diffraction efficiencies of the backward-diffracted (region 1) and forward-diffracted (region 3) waves in the directions $\mathbf{k}_{1i}$ and $\mathbf{k}_{3i}$, respectively.

Power conservation requires that for lossless phase gratings the sum of the efficiencies for all the propagating waves be unity. That is,

$$\sum_i (\mathrm{DE}_{1i} + \mathrm{DE}_{3i}) = 1. \tag{43}$$

It is important to note that Eq. (43) is satisfied for lossless phase gratings independently of the number of waves included in the analysis. Thus it sums to unity for any number of space harmonics retained, independently of whether the corresponding fields outside the grating are propagating or evanescent. Any significant deviation in the sum would indicate the presence of round-off errors in the numerical calculations. However, for all the calculations performed, the deviation was of the order of 10^{-12} (when a CDC 760/730 computer was used). However, the accuracy of each individual order depends on the number of space harmonics retained in constructing the $\mathbf{A}$ matrix. In all the results presented, these errors are less than 10^{-8}. It should be noted that this level of accuracy greatly exceeds that usually presented in grating-diffraction calculations.

In summary, the algorithm used to solve this problem proceeds as follows: First, the coefficient matrix $\mathbf{A}$ is constructed. Second, the eigenvalues and the eigenvectors are calculated. Third, the system of linear equations [Eqs. (31)–(40)] (or a modified version of these) is constructed and solved for the R_i's and T_i's. Fourth, the diffraction efficiencies are calculated using Eqs. (41) and (42).

It is important to note that, if it is desired to calculate the diffraction efficiencies for another polarization (another value of ψ), the new diffracted fields $\mathbf{R}_i(\psi)$ and $\mathbf{T}_i(\psi)$ may be cal-

culated for the new polarization ψ by using the values of $\mathbf{R}_i$ and $\mathbf{T}_i$ for any two noncollinear polarizations (e.g., ψ_1 and ψ_2) and the following relationships:

$$\mathbf{R}_i(\psi) = [\sin(\psi_2 - \psi)\mathbf{R}_i(\psi_1) - \sin(\psi_1 - \psi)\mathbf{R}_i(\psi_2)]/\sin(\psi_2 - \psi_1), \quad (44)$$

$$\mathbf{T}_i(\psi) = [\sin(\psi_2 - \psi)\mathbf{T}_i(\psi_1) - \sin(\psi_1 - \psi)\mathbf{T}_i(\psi_2)]/\sin(\psi_2 - \psi_1). \quad (45)$$

However, the values of $\mathbf{R}_i(\psi_1)$, $\mathbf{R}_i(\psi_2)$, $\mathbf{T}_i(\psi_1)$, and $\mathbf{T}_i(\psi_2)$ must still be determined by using the complete three-dimensional vector theory as described in this paper. This does *not* imply that the problem can be decomposed into uncoupled TE and TM problems.

The above method of solution applies equally for incidence at a Bragg angle or for a general angle of incidence. Bragg incidence occurs when the quantity m given by

$$m \equiv (2/K^2)[k_{x0}K_x + \mathrm{Re}(k^2\hat{\varepsilon}_0 - k_{x0}^2 - k_y^2)^{1/2}K_z] \quad (46)$$

is an integer. This would then correspond to the mth Bragg condition.

RESULTS AND DISCUSSION

The rigorous three-dimensional vector coupled-wave analysis presented in this paper describes the diffraction by an arbitrarily oriented planar grating with slanted fringes. In general, the grating vector does not lie in the plane of incidence, and conical diffraction results. For the special case when $\delta = 0$, the grating vector does lie in the plane of incidence, and it is possible to compare the results from the present vector theory with previously published rigorous scalar theory results.[8,9] Example results for $\delta = 0°$ are shown in Fig. 4 for an unslanted phase grating with a grating period equal to the wavelength of light in the medium. The normalized electric field and the diffraction efficiency of the first-order ($i = +1$) forward-diffracted wave are shown. When $\psi = 90°$ or $\psi = 0°$, the incident wave has TE or TM polarization, respectively. For $\psi = 90°$, the incident TE wave is coupled only to TE waves [such as the $i = +1$ diffracted TE wave shown in Fig. 4(a)]. However, there is no coupling of the incident TE wave to TM waves [see Fig. 4(b)]. Likewise, if the incident wave has TM polarization ($\psi = 0°$), it is coupled only to TM waves. There

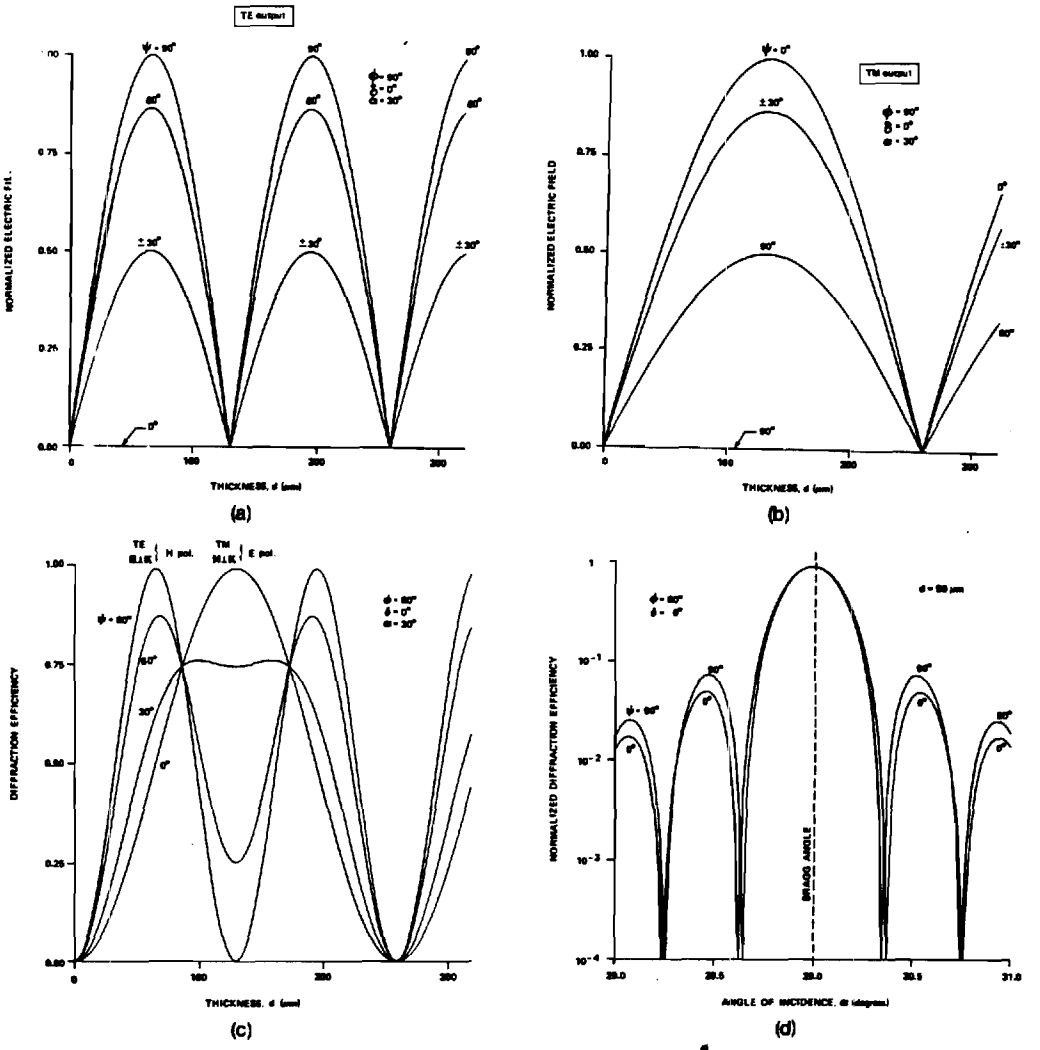


Fig. 4. Characteristics of first-order ($i = +1$) forward-diffracted wave for an unslanted-fringe grating ($\phi = 90°$) with a plane wave of wavelength $\lambda = 0.500\ \mu m$ incident at the first Bragg angle ($\alpha = 30°$) with the grating vector lying on the plane of incidence ($\delta = 0°$). The grating is lossless and has an average relative permittivity of $\varepsilon_0 = 2.25$ and a relative permittivity modulation of $\varepsilon_1 = 0.01$. The relative permittivity outside the grating region is the same as the average relative permittivity of the grating. (a) TE component of the normalized diffracted electric-field amplitude for various incident linear polarizations. (b) TM component of the normalized diffracted electric field. (c) Diffraction efficiency of diffracted wave. (d) Angular selectivity for a grating with a thickness of $d = 50\ \mu m$.

is no coupling of TM waves to TE waves in this $\delta = 0$ case. For the intermediate linear polarizations of the incident plane wave, the electric field can, therefore, be decomposed into TE and TM components, and the diffraction of these components may be treated entirely separately [see Eqs. (20)–(23) with $k_y = 0$]. The output diffracted field may then be obtained by vector addition of the individual diffracted TE and TM components. The resultant diffraction efficiencies are shown in Fig. 4(c). Since the conversion of incident TE to diffracted TE waves and the conversion of incident TM to diffracted TM waves have different coupling strengths, the resulting TE and TM diffracted wave amplitudes change at different rates with respect to grating thickness [compare Figs. 4(a) and 4(b)]. This out-of-phase behavior (with respect to grating thickness) causes the total normalized diffracted field to be less than unity, and thus the diffraction efficiency does not reach 100% with increasing thickness [see Fig. 4(c)]. However, for an incident polarization that is purely TE (and thus **E** is perpendicular to **K** for $\delta = 0°$) or purely TM (**H** perpendicular to **K** for $\delta = 0°$), the resultant diffraction efficiency will approach 100% with increasing thickness [see Fig. 4(c)]. For a grating of 50-μm thickness, the diffraction efficiency normalized to the value at the Bragg angle (30°) is shown as a function of angle of incidence in Fig. 4(d). This general form of angular selectivity is well known for thick gratings.[4] The angular width of the central angular-selectivity lobe is wider for E-mode polarization ($\psi = 0°$) than for H-mode polarization because the smaller coupling strength of that polarization produces less dephasing for a given angular deviation from the Bragg angle. All the numerical vector coupled-wave analysis calculations in Fig. 4 have been repeated using scalar rigorous coupled-wave analysis based on solving scalar-wave equations[8,9] for the individual TE and TM components. This method of analysis duplicates the results shown in Fig. 4.

Example results for a general-transmission grating when the grating vector does not lie in the plane of incidence are shown in Fig. 5. The grating and the wavelength are the same as those in Fig. 4. However, the plane of incidence is now inclined to $\delta = 30°$. For a general plane-of-incidence inclination angle δ, the Bragg condition [Eq. (46)] may be rewritten for a lossless grating as

$$m \equiv [2(\varepsilon_1)^{1/2}\Lambda/\lambda][\sin\alpha \sin\phi \cos\delta + (\varepsilon_0/\varepsilon_I - \sin^2\alpha)^{1/2} \cos\phi]. \quad (47)$$

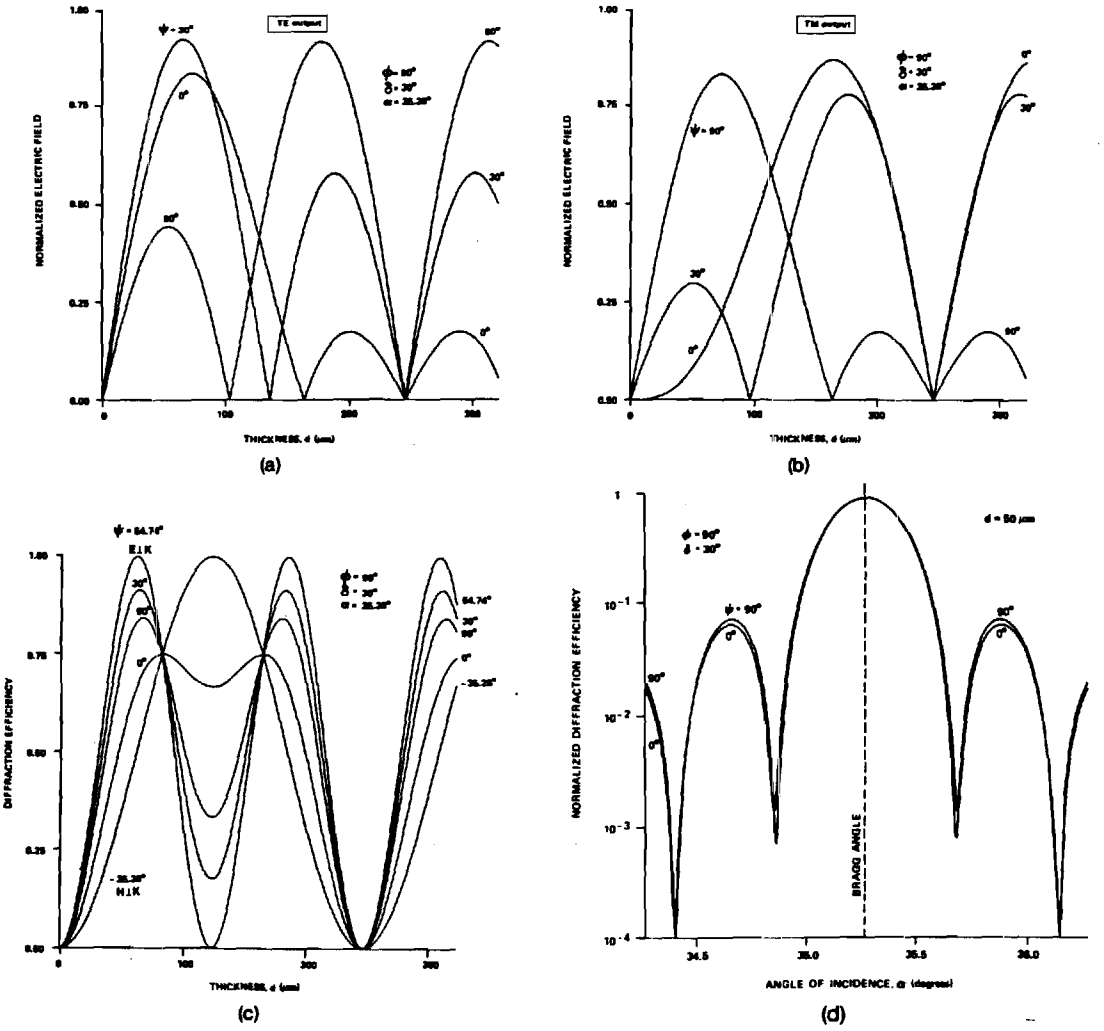


Fig. 5. Characteristics of first-order ($i = +1$) forward-diffracted wave for the same unslanted-fringe grating ($\phi = 90°$) as in Fig. 4 with a plane wave of wavelength of $\lambda = 0.500$ μm incident at the first Bragg angle ($\alpha = 35.26°$) with the plane of incidence inclined at an angle of $\delta = 30°$. The grating vector is therefore not in the plane of incidence. (a) TE component of the normalized diffracted electric field for various incident linear polarizations. (b) TM component of the normalized diffracted electric field. (c) Diffraction efficiency of diffracted wave. (d) Angular selectivity for a grating with a thickness of $d = 50$ μm.

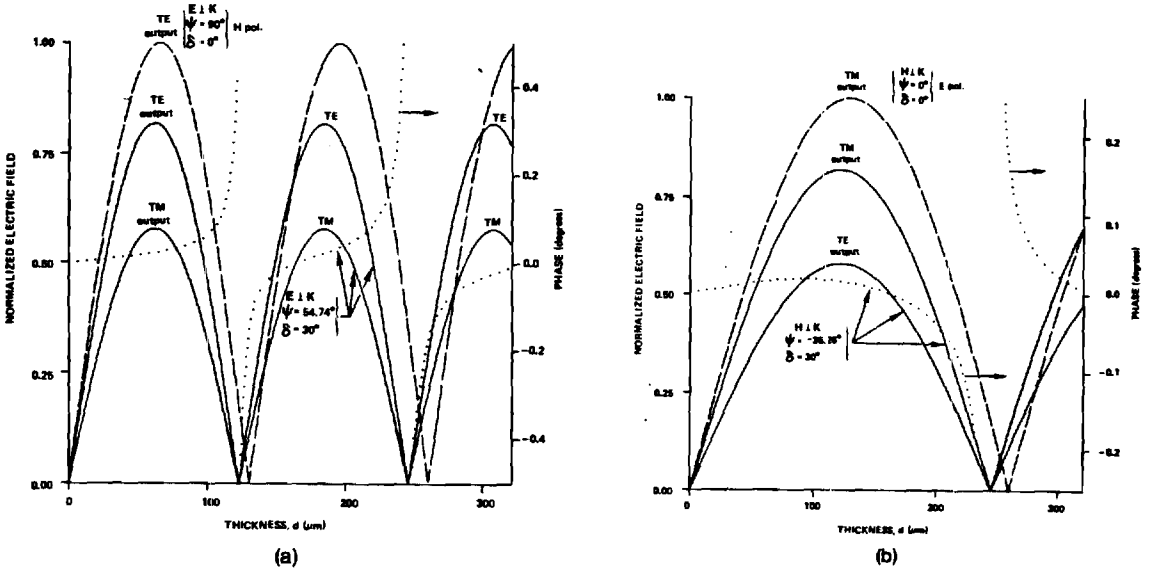


Fig. 6. Normalized diffracted electric-field amplitude for the same grating as in Figs. 4 and 5 for a grating vector both in plane of incidence ($\delta = 0°$) and out of plane of incidence ($\delta = 30°$) for (a) incident electric field perpendicular to the grating vector and (b) incident magnetic field perpendicular to grating vector. The solid and dashed lines are the normalized electric fields. The dotted line is the phase.

Solving this equation for the first Bragg angle ($m = 1$) gives $\alpha = 35.26°$. For an incident TE wave ($\psi = 90°$), coupling is now observed to both the TE diffracted wave *and* the TM diffracted wave [Figs. 5(a) and 5(b)]. Likewise, an incident TM wave ($\psi = 0°$) is coupled to both the TM diffracted wave *and* the TE diffracted wave.

Whether or not the grating vector lies in the plane of incidence, *it is shown that the diffraction efficiency can approach 100% in a lossless grating if the incident electric field (or magnetic field) is perpendicular to the grating vector* [see Figs. 4(c) and 5(c)]. The condition for **E** to be perpendicular to **K** is that $\mathbf{E} \cdot \mathbf{K} = 0$, and this may be written as

$$\tan \psi = \cos \alpha \cot \delta - \sin \alpha \csc \delta \cot \phi. \tag{48}$$

For the case represented by Fig. 5, this gives $\psi = 54.74°$. For **H** to be perpendicular to **K**, the value of ψ is changed by 90° from the value given by Eq. (48). *For the cases of incident* **E** *perpendicular to* **K** *(or incident* **H** *perpendicular to* **K**), *it is shown that the diffraction-efficiency minima in the angular-selectivity curves approach zero.* In Fig. 4(d), the angular-selectivity minima are nulls because the incident TE and TM ($\psi = 90°$ and $\psi = 0°$) waves have electric and magnetic fields, respectively, perpendicular to the grating vector. In Fig. 5(d), the incident TE and TM waves no longer have electric and magnetic fields perpendicular to **K**, and so nonzero minima occur in the angular-selectivity curves. This is due to different dephasing rates with changing angle of incidence for the resultant TE and TM components of the diffracted field.

Figure 6(a) shows the diffracted electric field for the case of **E** perpendicular to **K** for the grating of Figs. 4 and 5 ($\delta = 0°$ and $\delta = 30°$). Likewise, Fig. 6(b) shows the diffracted electric field for the case of **H** perpendicular to **K**. The grating for the four cases depicted in Fig. 6 is the same grating as in Figs. 4 and 5; only the incident direction and the incident polarization are changed. When the grating vector is in the plane of incidence $\delta = 0°$, the diffracted fields for incident **E** perpendicular to **K** and incident **H** perpendicular to **K** are completely decoupled, as was stated before. For the case of the grating vector not in the plane of incidence ($\delta = 30°$), incident **E** perpendicular to **K** results in TE and TM diffracted field components that oscillate with increasing thickness *with the same period*, as shown in Fig. 6(a). Because of this synchronism with grating thickness, the diffraction efficiency approaches 100% with increasing thickness [Fig. 5(c)]. This is in contrast to the TE and TM diffracted components of the other incident polarizations ($\psi = 0°, 30°, 90°$) depicted in Figs. 5(a) and 5(b) for $\delta = 30°$. Likewise, incident **H** perpendicular to **K** also results in TE and TM diffracted field components that are in synchronism with grating thickness as shown in Fig. 6(b), and this leads to essentially 100% diffraction efficiency, as shown in Fig. 5(c). Since there is less incident-wave to diffracted-wave coupling for incident **H** perpendicular to **K**, a larger grating thickness is required to achieve the same diffraction efficiency. The phase difference between the TE and TM components (Fig. 6, dotted lines) represents the de-

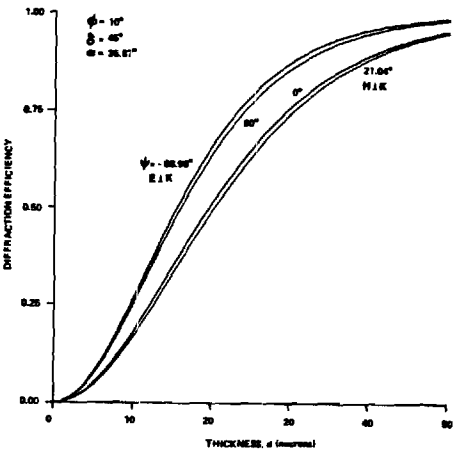


Fig. 7. Diffraction efficiency of first-order ($i = +1$) backward-diffracted wave for a slanted-fringe ($\phi = 10°$) grating with a plane wave of wavelength of $\lambda = 0.500\ \mu m$ incident at the first Bragg angle ($\alpha = 25.87°$) with the plane of incidence inclined at an angle of $\delta = 45°$. The grating vector is therefore not in the plane of incidence. The grating is lossless and has an average relative permittivity of $\varepsilon_0 = 2.25$ and a relative permittivity modulation of 0.01.

gree of ellipticity of the output polarization. For the case shown, the ellipticity is relatively small because of the average relative permittivity's being the same in all three regions. If these dielectric constants differed significantly in the three regions, the ellipticity would be correspondingly greater.

Figure 7 shows the diffraction efficiency of the first-order ($i = +1$) backward-diffracted wave that is due to a reflection grating ($\phi = 10°$). The grating vector does not lie in the plane of incidence ($\delta = 45°$). In this case the diffraction efficiencies for all linear polarization angles ψ approach 100% with increasing thickness. However, as before, *the coupling between incident and diffracted waves is greatest for the incident electric field perpendicular to the grating vector.* Thus the diffraction efficiency increases most rapidly with increasing thickness for that case.

The vector diffraction theory described in this paper can also be applied to surface-relief (corrugated) gratings by using the method of decomposition into planar gratings described in Ref. 31.

SUMMARY

A rigorous three-dimensional vector coupled-wave analysis of diffraction by a slanted-fringe grating has been presented for an arbitrary angle of incidence in three dimensions. The method applies to any sinusoidal or nonsinusoidal complex permittivity (amplitude and/or phase) grating and any linear polarization. Since no physical approximations are made, any arbitrary level of numerical accuracy may be achieved. It has been shown that the TE and TM components of an incident plane wave are each coupled to all the TE and TM components of all the forward-diffracted waves and backward-diffracted waves. For a general Bragg angle of incidence in three dimensions, it is shown that the diffraction efficiency can approach 100% for a lossless grating if either the incident electric field or the incident magnetic field is perpendicular to the grating vector. Maximum coupling between incident and diffracted waves is shown to occur when the incident electric field is perpendicular to the grating vector. In general, the diffracted waves are elliptically polarized. Even though it is possible to calculate the fields for any polarization (ψ) given the fields for two orthogonal polarizations, these orthogonally polarized fields must be first calculated from a general three-dimensional vector theory, such as that presented here, because of the inherent coupling between these fields.

ACKNOWLEDGMENT

This research was sponsored in part by the National Science Foundation and the Joint Services Electronics Program.

REFERENCES

1. R. R. Aggrawal, "Diffraction of light by ultrasonic waves," Proc. Indian Acad. Sci. **31**, 417–426 (1950).
2. P. Phariseau, "On the diffraction of light by progressive supersonic waves," Proc. Indian Acad. Sci. Sect. A **44**, 165–170 (1965).
3. W. R. Klein and B. D. Cook, "Unified approach to ultrasonic light diffraction," IEEE Trans. Sonics Ultrason. **SU-14**, 123–134 (1967).
4. H. Kogelnik, "Coupled wave theory for thick hologram gratings," Bell Syst. Tech. J. **48**, 2909–2947 (1969).
5. G. L. Fillmore and R. F. Tynan, "Sensitometric characteristics of hardened dichromated-gelatin films," J. Opt. Soc. Am. **61**, 199–203 (1971).
6. J. A. Kong, "Second-order coupled-mode equations for spatially periodic media," J. Opt. Soc. Am. **67**, 825–829 (1977).
7. R. Magnusson and T. K. Gaylord, "Analysis of multiwave diffraction by thick gratings," J. Opt. Soc. Am. **67**, 1165–1170 (1977).
8. M. G. Moharam and T. K. Gaylord, "Rigorous coupled-wave analysis of planar-grating diffraction," J. Opt. Soc. Am. **71**, 811–818 (1981).
9. M. G. Moharam and T. K. Gaylord, "Rigorous coupled-wave analysis of grating diffraction—*E*-mode polarization and losses," J. Opt. Soc. Am. **73**, 451–455 (1983).
10. T. Tamir, H. C. Wang, and A. A. Oliner, "Wave propagation in sinusoidally stratified dielectric media," IEEE Trans. Microwave Theory Tech. **MTT-12**, 323–335 (1964).
11. T. Tamir and H. C. Wang, "Scattering of electromagnetic waves by a sinusoidally stratified half-space: I. Formal solution and analysis approximations," Can. J. Phys. **44**, 2073–2094 (1966).
12. T. Tamir, "Scattering of electromagnetic waves by a sinusoidally stratified half-space: II. Diffraction aspects at the Rayleigh and Bragg wavelengths," Can. J. Phys. **44**, 2461–2494 (1966).
13. C. B. Burckhardt, "Diffraction of a plane wave at a sinusoidally stratified dielectric grating," J. Opt. Soc. Am. **56**, 1502–1509 (1966).
14. L. Bergstein and D. Kermisch, "Image storage and reconstruction in volume holography," Proc. Symp. Modern Opt. **17**, 655–680 (1967).
15. R. S. Chu and T. Tamir, "Guided wave theory of light diffraction by acoustic microwaves," IEEE Trans. Microwave Theory Tech. **MTT-18**, 486–504 (1970).
16. R. S. Chu and T. Tamir, "Wave propagation and dispersion in space–time periodic media," Proc. IEE **119**, 797–806 (1972).
17. F. G. Kaspar, "Diffraction by thick periodically stratified gratings with complex dielectric constant," J. Opt. Soc. Am. **63**, 37–45 (1973).
18. S. T. Peng, T. Tamir, and H. L. Bertoni, "Theory of periodic dielectric wavelengths," IEEE Trans. Microwave Theory Tech. **MTT-23**, 123–133 (1975).
19. R. S. Chu and J. A. Kong, "Modal theory of spatially periodic media," IEEE Trans. Microwave Theory Tech. **MTT-25**, 18–24 (1977).
20. M. Neviere, R. Petit, and M. Cadilhac, "About the theory of optical grating coupler-waveguide systems," Opt. Commun. **8**, 113–117 (1973).
21. M. Neviere, P. Vincent, R. Petit, and M. Cadilhac, "Systematic study of resonances of holographic thin film couplers," Opt. Commun. **9**, 48–53 (1973).
22. M. Neviere, P. Vincent, and R. Petit, "Sur la théorie du réseau conducteur et ses applications a l'optique," Nouv. Rev. Opt. **5**, 65–77 (1974).
23. K. C. Chang, V. Shah, and T. Tamir, "Scattering and guiding of waves by dielectric gratings with arbitrary profiles," J. Opt. Soc. Am. **70**, 804–813 (1980).
24. K. Knop, "Rigorous diffraction theory for transmission phase gratings with deep rectangular grooves," J. Opt. Soc. Am. **68**, 1206–1210 (1978).
25. D. Maystre, "Integral methods" in *Electromagnetic Theory of Gratings*, R. Petit, ed. (Springer-Verlag, Berlin, 1980).
26. S. L. Chuang and J. A. Kong, "Wave scattering from a periodic dielectric surface for a general angle of incidence," Radio Sci. **17**, 545–557 (1982).
27. M. G. Moharam and T. K. Gaylord, "Chain-matrix analysis of arbitrary-thickness dielectric reflection gratings," J. Opt. Soc. Am. **72**, 187–190 (1982).
28. T. K. Gaylord and M. G. Moharam, "Planar dielectric grating diffraction theories," Appl. Phys. B **28**, 1–14 (1982).
29. E.g., program EIGRF from the International Mathematics and Statistics Library, Houston, Texas.
30. E.g., B. Carnahan, H. A. Luther, and J. O. Wilkes, *Applied Numerical Methods* (Wiley, New York, 1969).
31. M. G. Moharam and T. K. Gaylord, "Diffraction analysis of dielectric surface-relief gratings," J. Opt. Soc. Am. **72**, 1385–1392 (1982).

Reprinted from Applied Optics, Vol. 23, page 3214, September 15, 1984

Diffraction characteristics of photoresist surface-relief gratings

M. G. Moharam, T. K. Gaylord, G. T. Sincerbox, H. Werlich, and B. Yung

Theoretical results from rigorous coupled-wave analysis are compared with experimental diffraction characteristics for holographically formed dielectric photoresist surface-relief gratings with deep grooves (greater than a grating period) and high diffraction efficiency (>85%). The angular selectivity (at a fixed wavelength) and the wavelength selectivity (at a fixed angle of incidence) are presented for both TE and TM incident polarizations. Modeling the gratings as a surface-relief modulated half-space and using rigorous coupled-wave analysis are shown to produce good general agreement with the experimentally measured diffraction characteristics.

I. Introduction

Dielectric surface-relief (or corrugated) gratings are of considerable interest due to their applications in quantum electronics, integrated optics, spectroscopy, and holography. Surface-relief gratings with high spatial frequencies (in excess of 2000 lp/mm) can be fabricated in photoresist using optical holography. These holographic optical elements (HOE) may be compact and lightweight compared to refractive optics, and they may be easy to replicate and thus are potentially low in cost. Example applications include waveguide couplers, spectral filters, wavelength multiplexers and dimultiplexers, scanners (such as supermarket point-of-sale laser scanners), and holographic beam combiners (such as in head-up displays).

Photoresist surface-relief gratings, when used in the transmission configuration (using forward-diffracted waves), have been traditionally low in diffraction efficiency (<30%).[1-3] However, dielectric surface-relief gratings have been analyzed recently by Moharam and Gaylord[4] using rigorous coupled-wave theory, and theoretical maximum diffraction efficiencies of >85% were shown to be possible. Recent experimental work of Enger and Case[5] and Werlich *et al.*[6] have shown that these high efficiencies are achievable in practice.

In this paper, the rigorous coupled-wave analysis is applied to photoresist surface-relief gratings with deep grooves fabricated by IBM.[6] This study compares theoretically calculated and experimentally measured angular selectivities (angle of incidence dependence of the diffraction efficiency at a fixed wavelength) and wavelength selectivities (wavelength dependence of the diffraction efficiency for a fixed angle of incidence) of these gratings for both TE and TM polarizations. Good general agreement is found for both the magnitude and functional form of the diffraction efficiencies.

II. Grating Diffraction Geometry

Diffraction by a photoresist surface-relief grating is shown schematically in Fig. 1. Grooves in the photoresist are periodically spaced with period Λ and groove depth d. An electromagnetic plane wave in air is incident upon the grating producing both forward-diffracted and backward-diffracted waves. The angles of diffraction of the forward-diffracted (transmitted) waves are given by the grating equation (from the phase-matching requirement)

$$i\lambda = \Lambda(\sin\theta' + n \sin\theta_i), \tag{1}$$

where i is the diffracted order, λ is the free-space wavelength, θ' is the angle of incidence in air, n is the refractive index of the photoresist, and θ_i is the angle of diffraction in the photoresist of the ith-order forward-diffracted wave. For the case illustrated in Fig. 1 ($\lambda = 458$ nm, $\Lambda = 458$ nm, $\theta' = 30°$, $n = 1.64$), there are four transmitted orders ($i = -1$ to $+2$) and two reflected orders ($i = 0, +1$). For other wavelengths and other angles of incidence, some of these orders may become evanescent (cut off), and this can affect the diffraction efficiency of the remaining propagating orders as will be shown below.

M. G. Moharam and T. K. Gaylord are with Georgia Institute of Technology, School of Electrical Engineering, Atlanta, Georgia 30332; the other authors are with IBM Research Laboratory, 5600 Cottle Road, San Jose, California 95193.

Received 14 February 1984.
0003-6935/84/183214-07$02.00/0.

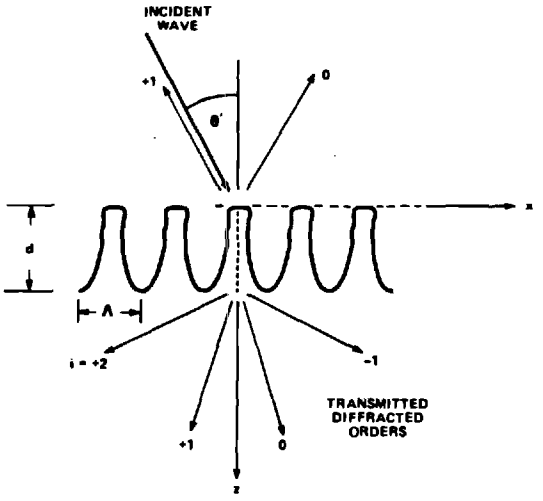


Fig. 1. Geometry of photoresist surface-relief grating.

Fig. 2. Scanning electron micrograph of typical photoresist surface-relief grating.

If the wavelength and angle of incidence satisfy the Bragg condition

$$m\lambda = 2\Lambda \sin\theta', \tag{2}$$

where m is an integer, the diffraction efficiency is generally (but not always) maximized for the ith diffracted order, where $i = m$. The first-order ($i = +1$) diffraction efficiency is likewise generally larger than the efficiencies associated with the other orders. In the present work, the first-order diffraction efficiency is of primary interest.

The photoresist surface-relief gratings treated in this work were holographically recorded in Shipley AZ-1400 series photoresist that had been spincoated at 4000 rpm onto a glass substrate. The photoresist was exposed at $\lambda = 458$ nm with two beams at 30° on opposite sides of the normal. The resulting grating period was thus also 458 nm, corresponding to a spatial frequency of 2183 lp/mm. During recording, an absorptive backing plate was index matched to the back of the glass to prevent reflections. The exposure energy ranged from 140 to 160 mJ/cm^2. The photoresist was developed with Shipley AZ 351 (30%) for 25 sec. The short wavelength ($\lambda = 458$ nm) at 30° incidence allows the gratings to be reconstructed with an ordinary He–Ne laser ($\lambda = 633$ nm) near $\theta' = 45°$ for applications requiring bow-free scanning.[7] A scanning-electron micrograph of a typical photoresist grating is shown in Fig. 2.

III. Rigorous Coupled-Wave Analysis

The rigorous coupled-wave analysis of grating diffraction was first applied to planar (volume) gratings.[8] In these gratings, the refractive index and/or optical absorption vary periodically between the two parallel planar surfaces of the grating. In this method, the field inside the grating is expanded in terms of space-harmonic components that have variable amplitudes in the thickness direction (z) of the grating. This field expansion together with the Floquet condition (due to the periodic nature of the structure) are then substituted into the appropriate (TE or TM polarization) wave equation, and an infinite set of coupled-wave equations is formed. Using a state space representation, this infinite set of second-order equations is converted to a doubly infinite set of first-order equations. The space-harmonic amplitudes are then solved for in terms of the eigenvalues and eigenvectors of the differential equation coefficient matrix. Applying boundary conditions (continuity of the tangential components of E and H across the boundaries), a set of linear equations is formed. Truncating this set of equations so that an arbitrary level of accuracy is achieved, the amplitudes of the propagating diffracted orders outside of the planar grating may then be determined. From these amplitudes the diffraction efficiencies may be directly calculated. None of the common approximations (neglect of second derivatives, neglect of boundary effects, neglect of higher-order waves, neglect of dephasing from the Bragg angle, or small grating modulation) are used in this analysis. The method is rigorous, and any specified level of accuracy can be obtained.

Rigorous coupled-wave analysis has also been applied to surface-relief gratings.[4] In this case, the surface-relief grating is divided into a large number of thin layers parallel to the surface. Each thin layer grating is analyzed using the state-variables method described above, and then by applying the boundary conditions to the boundaries of each thin layer, it is possible to obtain the forward- and backward-diffracted wave amplitudes. Typically 50–100 layers are used. Details of the rigorous coupled-wave approach applied to surface-relief gratings are given in Ref. 4.

IV. Diffraction Characteristics

Virtually all applications of gratings require knowledge of the maximum diffraction efficiency of the grating. However, in addition, practical applications such as those described above require knowledge of how the diffraction efficiency varies with (1) angle of incidence, (2) polarization, and (3) wavelength. These detailed characteristics are treated for photoresist

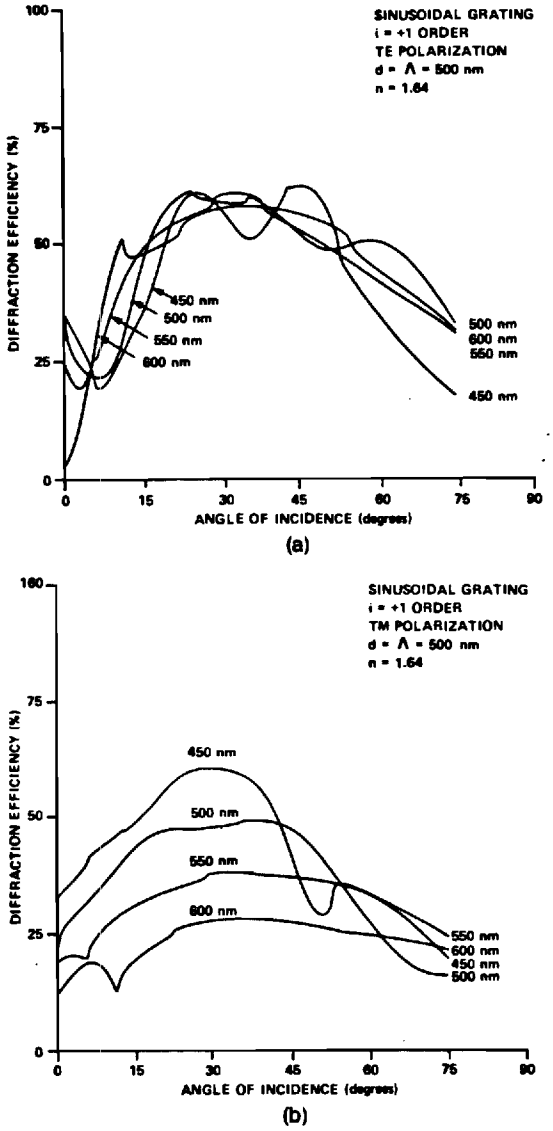


Fig. 3. Calculated angular dependence of the first-order ($i = +1$) transmitted diffraction efficiency with wavelength as a parameter for a sinusoidal surface-relief grating for (a) TE (or s) polarization, and (b) TM (or p) polarization. Rapid variations are observed when the electric field is parallel to the grooves (TE polarization).

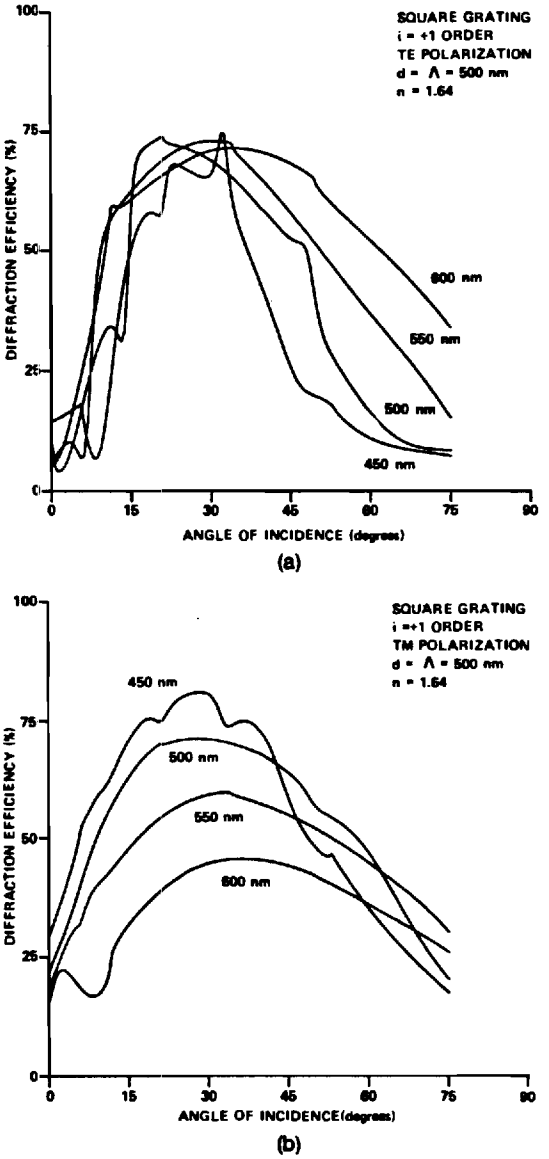
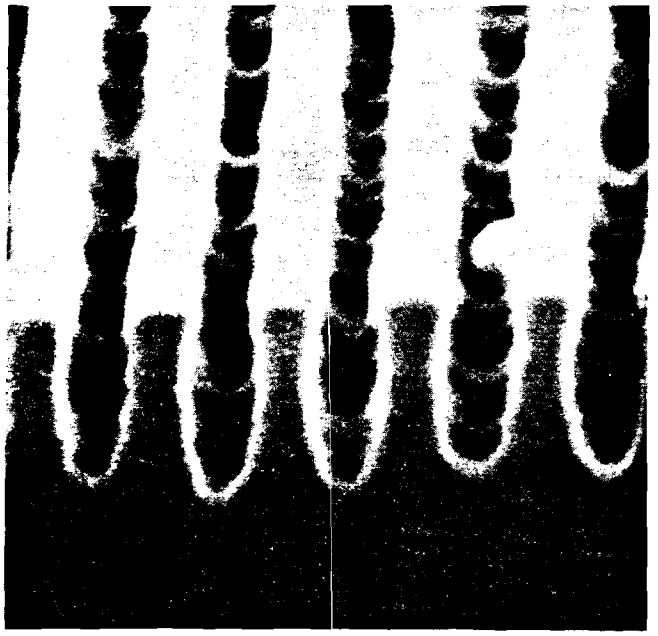


Fig. 4. Calculated angular dependence of the first-order ($i = +1$) transmitted diffraction efficiency with wavelength as a parameter for a square-wave surface-relief grating for (a) TE polarization and (b) TM polarization.

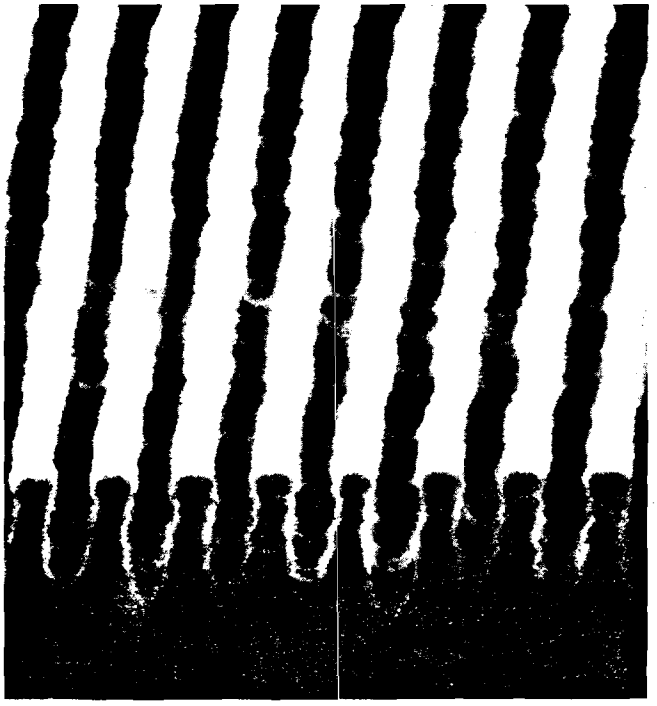
surface-relief gratings in this section, and both theoretical and experimental results are given.

Before modeling photoresist gratings using the surface-relief profiles as obtained from scanning electron micrographs, it is prudent to analyze some elementary profiles and to observe the general features of their diffraction characteristics. A photoresist grating with a sinusoidal surface profile has been analyzed using rigorous coupled-wave theory, and the dependence of the diffraction efficiency on angle of incidence in air is shown in Fig. 3 with polarization and wavelength as parameters. In this case the groove depth and the grating period are both equal to 500 nm. The photoresist is assumed to have a refractive index of 1.64. The diffraction efficiency of the first-order ($i = +1$) wave is shown for the incident electric field perpendicular to the plane of incidence (TE or s polarization) in Fig. 3(a) and for the electric field in the plane of incidence (TM or p polarization) in Fig. 3(b).

Similarly, a photoresist grating with a square-wave profile has been analyzed with rigorous theory, and the diffraction results are presented in Fig. 4. The groove depth is again equal to the grating period (500 nm). For a given wavelength and angle of incidence, the square-wave grating has a somewhat higher diffraction efficiency than does the sinusoidal grating. This, of course, is for the case of equal groove depths. If the groove depth can be chosen to maximize the diffraction efficiency, that sinusoidal surface profile is capable of a higher maximum efficiency than is a square-wave grating.[4] For both the sinusoidal and square-wave gratings, Figs. 3 and 4 show that the diffraction effi-



(a)



(b)

Fig. 5. Scanning electron micrographs of photoresist surface-relief gratings (a) 7 and (b) 8 analyzed in this work. All gratings were fabricated by IBM.

ciency varies more rapidly with changes in angle of incidence and with changes in wavelength for the TE polarization than for the TM polarization. When the incident electric field is parallel to the surface-relief grooves (TE polarization), small changes in angle or wavelength can produce relatively large changes in the diffraction efficiency. When the incident electric field is perpendicular to the surface-relief grooves (TM polarization), the diffraction efficiency varies much more smoothly with changes in angle or wavelength.

The angular and wavelength selectivities of a grating are measures of the sensitivity of the diffraction to changes in the angle of incidence (with the wavelength fixed) and to changes in the wavelength (with the angle of incidence fixed). The angular selectivity $\Delta\theta$ for a given wavelength may be defined as the full angular deviation about the first Bragg angle ($m = 1$) that causes a reduction in the diffraction efficiency to a specified fraction of its value at the Bragg angle. The angular selectivity is given by

$$\Delta\theta' = \theta'^{+} - \theta'^{-}, \tag{3}$$

where θ'^{+} and θ'^{-} are the angles of incidence, greater than and less than the Bragg angle, respectively, at which the diffraction efficiency has dropped to the specified fraction of the on-Bragg angle value. The wavelength selectivity, $\Delta\lambda$ for a given angle of incidence may likewise be defined as the full wavelength deviation about the first Bragg wavelength ($m = 1$) that causes a reduction in the diffraction efficiency to a specified fraction of its value at the Bragg wavelength. The wavelength selectivity is given by

$$\Delta\lambda = \lambda^{+} - \lambda^{-}, \tag{4}$$

where λ^{+} and λ^{-} are the wavelengths, greater than and less than the Bragg wavelength, respectively, at which the diffraction efficiency has dropped to the specified fraction of the Bragg wavelength value. The angular and wavelength selectivities may be calculated with the rigorous coupled-wave theory described in Sec. III. In addition, these values may be roughly estimated using the approximate two-wave first-order theory for planar gratings of Kogelnik.[9] From this theory, the upper and lower angles and wavelengths for TE polarized light are given by

$$\theta'^{\pm} = \sin^{-1}\left\{n\,\frac{\sin\theta_B \pm (\xi\Lambda/\pi d)[\cos^2\theta_B - (\xi\Lambda/\pi d)^2]^{1/2}}{1 + (\xi\Lambda/\pi d)^2}\right\}, \tag{5}$$

$$\lambda^{\pm} = \lambda_B \pm 2n\Lambda^2\xi\cos\theta_B/\pi d. \tag{6}$$

The quantity ξ is a dephasing parameter. If $\xi = 0$, there is no dephasing and $\theta^{\pm} = \theta_B$ and $\lambda^{\pm} = \lambda_B$ indicating $\Delta\theta = 0$ and $\Delta\lambda = 0$ (Bragg condition satisfied). For a grating with 75% diffraction efficiency at the first Bragg condition ($m = 1$), the dephasing parameter is $\xi = 0.5375$ for the diffraction efficiency to decrease to 90% of its value at the Bragg condition. For gratings with an external Bragg angle of incidence of 30°, $\lambda = 500$ nm, $\Lambda = 500$ nm, $d = 500$ nm, and $n = 1.64$ as represented in Figs. 3 and 4, these approximate relations predict $\Delta\theta' = 35.7°$ and $\Delta\lambda = 534.5$ nm.

Scanning electron photomicrographs of two actual photoresist surface-relief grating profiles fabricated and characterized at IBM are shown in Fig. 5. The profiles are neither sinusoidal nor square wave in shape but have some features of both. In addition, small local variations in the profiles are also apparent and are characteristic of photoresist gratings. For purposes of calculation, representative periods from the scanning

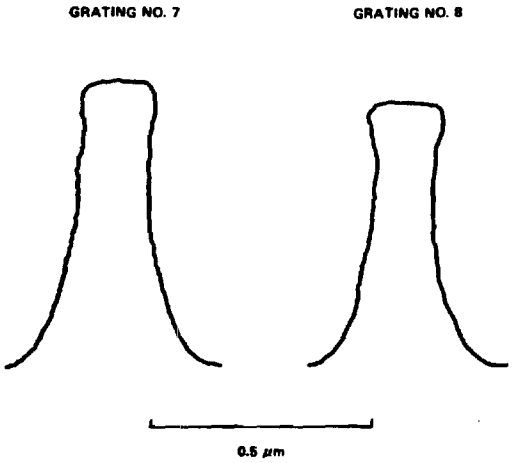


Fig. 6. Surface-relief profiles selected for the calculations. The groove depths are 0.65 and 0.59 μm for gratings 7 and 8, respectively.

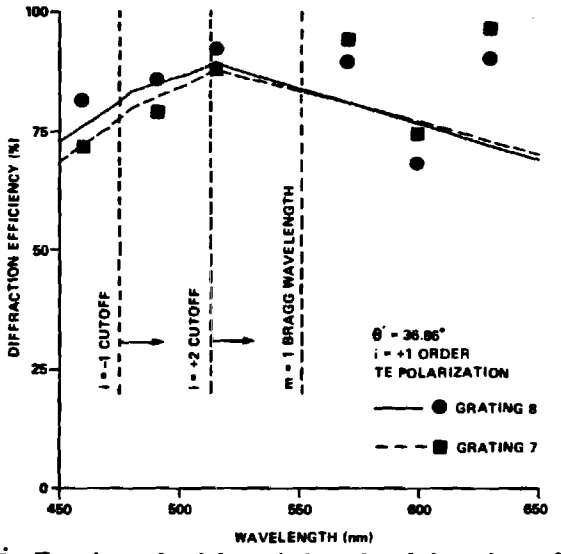


Fig. 7. Experimental and theoretical wavelength dependence of the diffraction efficiency for a fixed angle of incidence ($\theta' = 36.86°$).

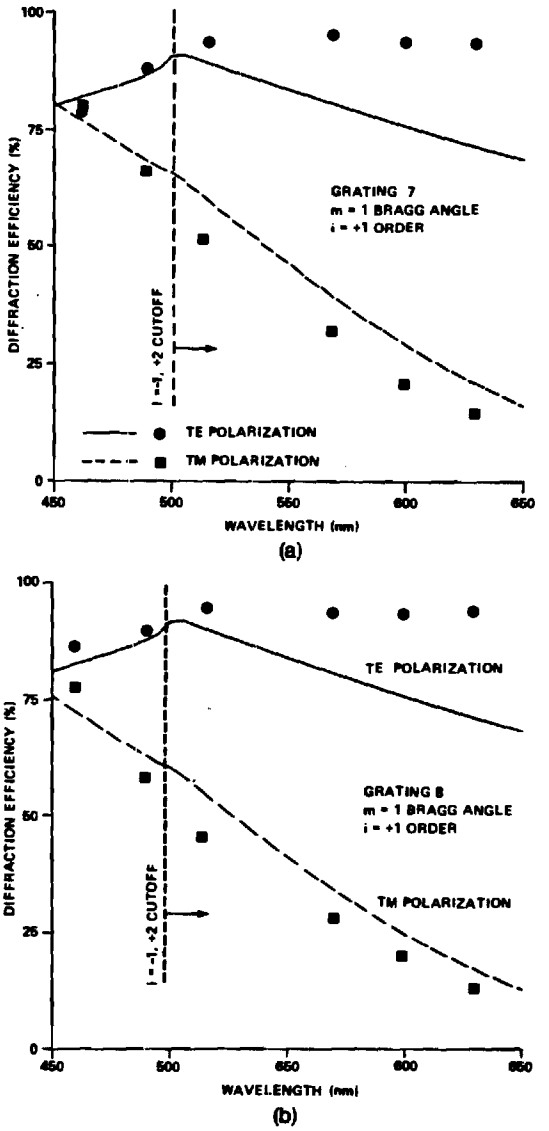


Fig. 8. Experimental and theoretical wavelength dependence of the diffraction efficiency with the angle of incidence continually adjusted to the first Bragg angle ($m = 1$) for (a) grating 7 and (b) grating 8.

electron micrographs were digitized and modeled by a stack of thin gratings to obtain a high-resolution replica of the actual surface profile. The modeled profile shapes for gratings 7 and 8 are shown in Fig. 6. The groove depths are 0.65 and 0.59 μm for gratings 7 and 8, respectively. Measured diffraction characteristics of these two gratings are shown as data points in Figs. 7–9. The experimental diffraction efficiencies shown represent the power in the first-order ($i = +1$) forward-diffracted beam divided by the sum of the $i = 0$ and $i = +1$ forward propagating powers.

Wavelength selectivity results are shown for both gratings in Fig. 7. The angle of incidence is fixed ($\theta' = 38.86°$), and when the wavelength is equal to 550 nm, the first ($m = 1$) Bragg condition is satisfied. Using the rigorous coupled-wave theory, the continuous curves shown in Figs. 7–9 were obtained. In these calculations, the grating is modeled as a surface-relief modulated half-space, and the effect of the back surface is neglected. In addition, the calculated diffraction efficiency is given by the diffracted power divided by the entire incident power. Thus, unlike the experimental diffraction efficiency definition, the power lost into the backward-traveling orders is taken into account. However, rigorous coupled-wave calculations of the backward-traveling orders as shown in Ref. 4 predict these powers to be small. For a wavelength of $\lambda = 450$ nm, there are four transmitted diffracted orders. This is represented by the case depicted in Fig. 1. As the wavelength increases, the transmitted diffracted orders spread apart in angle. As given by Eq. (1) the $i = -1$ forward-diffracted order becomes cut off (evanescent) at $\lambda = 476.4$ nm producing only three forward-diffracted orders. At $\lambda = 512.9$ nm, the $i = +2$ forward-diffracted order becomes cut off, and then there are only two forward propagating orders ($i = 0, +1$). Near a cut-off

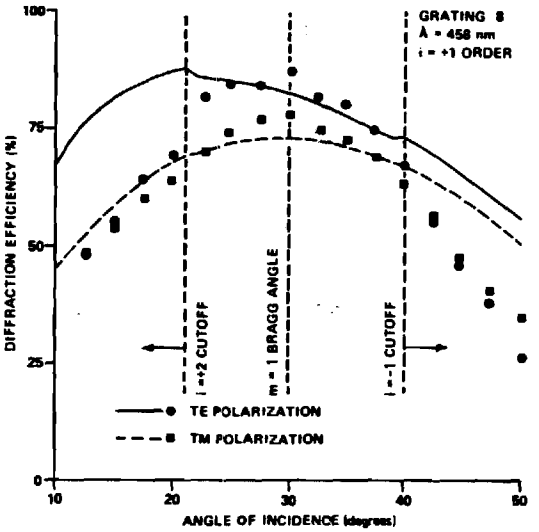


Fig. 9. Experimental and theoretical angle of incidence dependence of the diffraction efficiency for a fixed wavelength ($\lambda = 458$ nm) for TE and TM polarizations.

wavelength of a higher order, the form of the $i = +1$ diffraction efficiency generally changes, and this is a type of Wood's anomaly.[10] These changes in the $i = +1$ diffraction efficiency are apparent in Figs. 7–9. Further wavelength selectivity results are shown in Fig. 8. In this case, as the wavelength is changed, the angle of incidence is also changed so that the Bragg condition given by Eq. (2) is always satisfied with $m = 1$. In this situation both the $i = -1$ and $i = +2$ forward-diffracted orders become cut off at $\lambda = 549.5$ nm due to the symmetry imposed by always having the angle of incidence at the Bragg angle. The diffraction efficiencies for TE polarization are usually larger than those for TM polarization. The change in the form of the diffraction efficiency when higher-order waves become cut off is seen to be much more apparent for TE polarization than for TM polarization, and this is opposite of that generally observed for metallic reflection gratings.[10] As in Fig. 7, gratings 7 and 8 are again seen to produce similar results.

Angular selectivity results for grating 8 are shown in Fig. 9. In this case, the wavelength remains fixed at the recording wavelength $\lambda = 458$ nm, and the angle of incidence is varied. The Bragg condition with $m = 1$ is satisfied for an angle of incidence of 30°. Decreasing the angle of incidence to 21.1° causes the $i = +2$ wave to become cut off, and increasing it to 39.8° causes the $i = -1$ wave to become cut off. The diffraction efficiency for TE polarization is again generally larger than that for TM polarization. The effects of Wood's anomalies are clearly seen in the calculated TE polarization results but are barely visible in the calculated TM polarization diffraction efficiencies. The experimentally measured diffraction efficiencies decrease away from the Bragg angle more rapidly than is predicted by theory.

In grating diffraction, it is common to have the maximum diffraction efficiency occur when the Bragg condition is satisfied. However, for highly modulated gratings such as surface-relief gratings, this may not always occur. The calculated diffraction efficiencies given in Figs. 3(a) and (b), 4(a), 7, and 9 provide examples for both TE and TM polarizations of the diffraction efficiency increasing (rather than decreasing) as the angle of incidence or the wavelength is deviated from the value that satisfies the $m = 1$ Bragg condition.

V. Summary and Discussion

Diffraction efficiency results for photoresist surface-relief gratings fabricated by IBM have been calculated using rigorous coupled-wave theory and compared with experimentally obtained diffraction characteristics. The diffraction characteristics as a function of (1) angle of incidence, (2) polarization, and (3) wavelength have all been measured experimentally and calculated theoretically using a surface-relief modulated half-space model. Good general agreement is found between theory and experiment. Because the thicknesses are small compared with those of most planar-type gratings, the angular selectivities and wavelength selectivities are large, and thus the grating diffraction characteristics are relatively insensitive to variations in angle of incidence and wavelength when compared to thick planar gratings. Due to the surface-relief modulation as opposed to relatively small refractive-index modulations in planar gratings, these gratings are capable of high diffraction efficiencies (>85%).

The diffraction regimes associated with planar-surface (slab) dielectric gratings may be accurately delineated in terms of parameters ρ, Q', and γ as defined and discussed in Ref. 11. For Bragg regime diffraction as given by Kogelnik's theory[9] (one possible definition of thick grating behavior[11]), the required condition is $\rho \gg 1$. For Raman-Nath regime diffraction characteristics as given by the theory of Raman and Nath[12–15] (one possible definition of thin grating behavior[11]), the required condition is $Q'\gamma \ll 1$. Even though these grating parameters were developed for planar-surface slab gratings, they can be used to give an approximate indication of the diffraction behavior of the present photoresist surface-relief gratings. In spite of the fact that the photoresist surface-relief gratings are physically thin, their diffraction characteristics are not described by Raman-Nath thin grating theory since they are effectively very highly modulated. According to Raman-Nath theory, the maximum $i = +1$ diffraction efficiency is 33.8%. Apparently, by accident, the early experimentally measured values of diffraction efficiency[1–3] did not exceed 30% in agreement with Raman-Nath theory. However, for the surface-relief gratings such as those treated in this paper the corresponding grating parameters are $\rho \simeq 0.9$ and $Q'\gamma \simeq 6$. Since the condition $Q'\gamma \ll 1$ is clearly not met, Raman-Nath thin grating diffraction behavior should not be expected. Indeed for these values of ρ and $Q'\gamma$, it is apparent that approximate theories will be unable to give accurate results and that rigorous theory is definitely necessary.

It has recently been shown[16,17] that photoresist gratings can be transferred to glass or quartz substrates

using reactive ion etching. This eliminates the problem of potential damage to the photoresist grating and provides further motivation for using dielectric surface-relief gratings.

This work was sponsored in part by the Joint Services Electronics Program.

References

1. R. A. Bartolini, "Characteristics of Relief Phase Holograms Recorded in Photoresist," Appl. Opt. **13**, 129 (1974).
2. F. Iwata and J. Tsujiuchi, "Characteristics of a Photoresist Hologram and Its Replica," Appl. Opt. **13**, 1327 (1974).
3. R. Kurtz and R. Owen, "Holographic Recording Materials—A Review," Opt. Eng. **14**, 393 (1975).
4. M. G. Moharam and T. K. Gaylord, "Diffraction Analysis of Dielectric Surface-Relief Gratings," J. Opt. Soc. Am. **72**, 1385 (1982).
5. R. C. Enger and S. K. Case, "High-Frequency Holographic Transmission Gratings in Photoresist," J. Opt. Soc. Am. **73**, 1113 (1983).
6. H. Werlich, G. Sincerbox, and B. Yung, "Fabrication of High Efficiency Surface Relief Holograms," IBM Research Report RJ3912 (1983).
7. C. J. Kramer, "Hologon Laser Scanners for Nonimpact Printing," Proc. Soc. Photo-Opt. Instrum Eng. **390**, 165 (1982).
8. M. G. Moharam and T. K. Gaylord, "Rigorous Coupled-Wave Analysis of Planar-Grating Diffraction," J. Opt. Soc. Am. **71**, 811 (1981).
9. H. Kogelnik, "Coupled Wave Theory for Thick Hologram Gratings," Bell Syst. Tech. J. **48**, 2909 (1969).
10. For example, M. C. Hutley, *Diffraction Gratings* (Academic, London, 1982).
11. T. K. Gaylord and M. G. Moharam, "Thin and Thick Gratings: Terminology Clarification," Appl. Opt. **19**, 3271 (1981).
12. C. V. Raman and N. S. N. Nath, "The Diffraction of Light by High Frequency Sound Waves: Part I," Proc. Indian Acad. Sci. Sect. A **2**, 406 (1935).
13. C. V. Raman and N. S. N. Nath, "The Diffraction of Light by High Frequency Sound Waves: Part II," Proc. Indian Acad. Sci. Sect. A **2**, 413 (1935).
14. C. V. Raman and N. S. N. Nath, "The Diffraction of Light by High Frequency Sound Waves: Part III," Proc. Indian Acad. Sci. Sect. A **3**, 75 (1936).
15. C. V. Raman and N. S. N. Nath, "The Diffraction of Light by High Frequency Sound Waves: Part IV," Proc. Indian Acad. Sci. Sect A **3**, 119 (1936).
16. R. C. Enger and S. K. Case, "Optical Elements with Ultrahigh Spatial-Frequency Surface Corrugations," Appl. Opt. **22**, 3220 (1983).
17. M. Gale and K. Knop, *Surface Relief Images for Color Reproduction* (Focal Press, Woburn, Mass., 1980).